The Problem with Scorecards: How (and how not) to Measure the Cost-Effectiveness of Economic Sanctions

Richard Parker
University of Connecticut School of Law

Follow this and additional works at: http://digitalcommons.uconn.edu/law_papers

Recommended Citation
Parker, Richard, "The Problem with Scorecards: How (and how not) to Measure the Cost-Effectiveness of Economic Sanctions" (2000). Faculty Articles and Papers. 118.
http://digitalcommons.uconn.edu/law_papers/118
THE PROBLEM WITH SCORECARDS:
HOW (AND HOW NOT) TO MEASURE
THE COST-EFFECTIVENESS OF
ECONOMIC SANCTIONS

Richard W. Parker*

I. MEASURING AND REPORTING THE COST OF SANCTIONS
   TO THE U.S. ECONOMY ................................................................. 242
   A. Costing Methodology and Conclusions ...................................... 243
   B. Critique ................................................................................. 245
II. MEASURING THE EFFECTIVENESS OF TRADE SANCTIONS .......... 248
   A. Stories and Studies .............................................................. 249
   B. Scorecards ............................................................................ 251
   C. Critique of the Scorecard Methodology .................................. 258
      1. Lack of Control Group or Baseline ....................................... 259
      2. Selection Bias .................................................................... 260
      3. Subjectivity Problems ....................................................... 262
      4. Proxy Problems ................................................................... 264
      5. The Difficulty of Defining the Temporal Boundaries and Baselines ............................................. 265
      6. Model Mis-specification ....................................................... 268
      7. Problems of Endogeneity .................................................... 276
      8. Multicollinearity ............................................................... 278
      9. The Difficulty of Modeling Dynamic Processes by Static Proxies ....................................................... 279
      10. Invalid Cross-tabs Comparisons ......................................... 280
      11. Data Problems .................................................................... 281
      12. Invalid Extrapolation from Foreign Policy to Environmental Sanctions ............................................... 283
III. TOWARDS AN EMPIRICAL FOUNDATION FOR
    SANCTIONS POLICY ................................................................. 288
CONCLUSION ................................................................................. 295

* Professor of Law, University of Connecticut School of Law; Visiting Professor, Georgetown University Law Center, 2000; J.D. Yale Law School, 1985, D.Phil., Politics (International Relations), Oxford University, 1982; B.A., Princeton University, 1978. The author has previously served as Assistant General Counsel at the Office of the United States Trade Representative and Special Counsel to the Deputy Administrator of the U.S. Environmental Protection Agency. I would like to thank John J. Donohue III for helpful comments on an earlier draft. Special thanks to Peter Siegelman for valuable insights on econometric sources and methods as well as helpful comments on multiple drafts of this article.
The United States has long relied on economic sanctions for a simple reason: they seem to offer a middle path between violence and talk. This basic attraction has prompted the U.S. to use economic sanctions on over 200 occasions in this century for a wide variety of purposes ranging from weakening adversaries, toppling governments, and promoting human rights\(^1\) to opening foreign markets, protecting intellectual property,\(^2\) and conserving the global commons.\(^3\)

In recent years, however, the use of economic sanctions has come in for stricter scrutiny than ever before. One reason has been a resurgence

---

1. Although the U.S. is a frequent user of sanctions it is by no means the only user. Of the 116 post-World War I foreign policy sanctions episodes examined by Hufbauer, Schott and Elliott through 1990, 77 involved the United States either alone or in concert with allies. The remaining 39 cases involved exclusively non-US. sending states, and the authors admit they probably missed a number of cases involving foreign 'senders' of sanctions. GARY CLYDE HUFBAUER ET AL., ECONOMIC SANCTIONS RECONSIDERED: HISTORY AND CURRENT POLICY, 8–9 (2d ed. 1990) [hereinafter ECONOMIC SANCTIONS RECONSIDERED (2d ed.)]. See also HUFBAUER ET AL., ECONOMIC SANCTIONS RECONSIDERED: HISTORY AND CURRENT POLICY 13–20 (1st ed. 1985) [hereinafter ECONOMIC SANCTIONS RECONSIDERED (1st ed.)] (listing the 103 cases analyzed in that edition).

2. Here, the most frequent user has been the U.S. followed by the European Union. Since 1974, Section 301 of the Trade Act of 1974 has authorized the President to retaliate against foreign trade practices that infringe GATT rights or are otherwise 'unreasonable.' Bayard and Elliott report that the United States overtly invoked Section 301 of the Trade Act of 1974 in 89 separate cases over the period 1974–1992. Thomas O. Bayard & Kimberly A. Elliott, 'Aggressive Unilateralism' and Section 301: Market Opening or Market Closing?, 15 WORLD ECON. 685, 687 (1992) [hereinafter, Aggressive Unilateralism]; Alan O. Sykes, Constructive Unilateral Threats in International Commercial Relations: The Limit Case for Section 301, 23 L. & POL'Y INT'L. BUS. 263 (1992) [hereinafter Constructive Unilateral Threats].

3. About 20 multilateral agreements currently require or authorize nations to restrict imports or exports of goods or services to promote some conservation purpose. See WTO Secretariat, Trade Measures for Environmental Purposes Take Pursuant to Multilateral Environmental Agreements: Recent Developments, WTO Doc. PC/SCTE/W/3 (Oct. 13, 1994). The most significant of these trade-related MEAs are the Basel Convention on the Control of Trans-boundary Movements of Hazardous Waste (Basel Convention), the Convention on International Trade in Endangered Species, and the Montreal Protocol on Substances That Deplete the Ozone Layer (Montreal Protocol). In addition, the United States has, on occasion, used or threatened targeted trade restrictions to: (1) curb illegal trade in body parts of endangered rhinos and tigers (China and Taiwan); (2) deter illegal taking of depleted Atlantic bluefin tuna in violation of IATTC norms (Belize and Honduras); (3) enforce the IWC moratorium on commercial whaling (Norway, Russia, Iceland, Japan); (4) uphold General Assembly Resolutions banning the use of large scale driftnets on the high seas (Italy, Taiwan, Korea); (5) protect endangered sea turtles in the shrimp trawling process (a number of Latin American and Asian fleets); and (6) conserve dolphins in the Eastern Tropical Pacific tuna fishery (10–12 nations). Id. The European Union, a less frequent user, recently enacted a ban on import of fur from countries (such as the U.S. and Canada) that allow the use of leg-hold traps. For an excellent overview of cases involving the non-U.S. and foreign use of trade leverage for global environmental purposes both unilaterally and pursuant to multilateral agreements, see Robert E. Hudec, GATT Legal Restraints on the Use of Trade Measures against Foreign Environmental Practices, in 2 FAIR TRADE AND HARMONIZATION 95, 98–106 (Jagdish Bhagwati & Robert E. Hudec eds., 1996) [hereinafter GATT Legal Restraints].
in the use of sanctions in recent years.\textsuperscript{4} The main reason for the current preoccupation with sanctions reform, however, is a sustained media and lobbying initiative organized by the U.S. business community under the banner of “USA*Engage.”\textsuperscript{5} The campaign, begun in 1997, has already won a small deluge of newspaper editorials and stories, virtually all critical of sanctions.\textsuperscript{6} It has achieved the introduction of legislation that would impose procedural shackles on all new sanctions.\textsuperscript{7} Meanwhile, the movement appears already to have accomplished a significant change in the mental grid through which congressional leaders perceive sanctions: a long-standing congressional presumption in favor of sanctions appears to have been replaced—in the minds of prominent lawmakers—by a presumption against.\textsuperscript{8}


\textsuperscript{5} For an overview of this campaign and its message see USA Engage—About Us, (visited Feb. 12, 2000) <www.usaengage.org/backgroung/about.html>.


\textsuperscript{7} See, for example, \textit{The Sanctions Policy Reform Act} (S. 757), introduced March 24, 1999, which would establish a procedural rule requiring detailed analysis of the domestic and foreign economic impact of any proposed new sanction, the likely effectiveness of the measure if adopted, and available alternatives to sanctions. It would also sunset each and every sanction provision after two years unless it is affirmatively re-enacted by Congress, and give the President broad waiver authority to rescind sanctions determined to be too costly or ineffective. \textit{Sanctions Policy Reform Act}, S. 757, 106th Cong. § 1 (1999). \textit{See also} Enhancement of Trade, Security, and Human Rights through Sanctions Reform Act, H.R. 2708, 105th Cong. § 1 (1997).

\textsuperscript{8} As Richard Lawrence observed in the \textit{Journal of Commerce} on August 2, 1999, “Congress’ attitude toward sanctions appears much changed from only a short time ago. In September 1997, roughly 20 bills advocated tighter sanctions. Now, of roughly three dozen
Powering the USA*Engage campaign, intellectually, are two broad claims: (1) economic sanctions are costly to U.S. businesses, farmers and workers; and (2) economic sanctions are seldom effective and sometimes counter-productive. Hence the call for more rigorous prior analysis of sanctions proposals. In itself, of course, a call for closer scrutiny of the costs and benefits of sanctions would seem virtually unopposable. But the requirement does bring to the fore what turns out to be a very difficult question: how do we gauge the cost and effectiveness of economic sanctions, either ex ante or ex post? In other words, how do we know what we think we know about the cost and effectiveness of economic sanctions?

This Article looks "beneath the bridge" of sanctions law and policy to investigate these foundational questions. Part I will look briefly behind the currently prevailing estimate for the direct economic cost of high foreign policy export sanctions for the U.S. economy. It will demonstrate that the most widely reported aggregate cost estimate of $15-20 billion per year and 200,000 U.S. jobs lost is unsubstantiated. Moreover, the evidence is clear that environmental trade sanctions, i.e., import restrictions deployed for environmental purposes, have cost U.S. companies and workers virtually nothing. Trade sanctions may impose very significant costs on individual companies, and these costs may raise issues of fairness and the appropriateness of compensation. But there is no bills pending, most would rescind sanctions or reform the sanctions process." Sanctions Debate, supra note 6, at 8.


10. Although those trade policy sanctions aimed at promoting commercial interests are exempted from the proposed legislation, measures devoted to international environmental goals are tarred by the same brush as “high foreign policy” sanctions aimed at punishing foreign dictators, etc. See Sanctions Policy Reform Act, S. 757, at § 4(1)(G).


12. See infra note 39 and accompanying discussion. Sanctions may, of course, carry very high moral, humanitarian and political costs for both target and sending states. These costs are very important elements which I will address indirectly under the heading of “effectiveness.” See discussion infra Part II.B.5 and 6 (discussing the importance of political costs in sanctions impact assessments).
evidence to support the notion that the current regime of economic sanctions poses any significant threat to the U.S. economy as a whole.

Part II takes up the more significant issue of how we know what we think we know about the effectiveness of sanctions in achieving U.S. policy goals. By “effectiveness” I mean the contribution of sanctions to achieving desired changes in target state behavior—excluding any symbolic, expressive, or domestic political function.¹³ It will be seen that the case for the general ineffectiveness of sanctions is built on a combination of anecdotes, case studies, and multi-case “scorecards” of sanctions successes and failures. But anecdotes, while highly beneficial as specific illustrations, are too shallow and selective, analytically, to provide a conclusive foundation for sanctions policy. Case studies, while deeper and more nuanced, are narrow and difficult to digest.¹⁴ Too often the insights yielded by case studies are shrouded in detail and entombed within the volumes in which they first appeared. Scorecards, on the other hand, seem to capture the best of both approaches. They draw on in-depth case studies to explore the respective contribution of multiple factors (including sanctions) across a wide range of cases. Moreover, they often employ quantitative techniques which yield a numerical “success rate” for sanctions, as well as a finding of “significance” (or not) for each of any number of independent variables tested.¹⁵ This appealing combination of brevity, comprehensiveness, and quantitative rigor has made scorecards by far the most influential indices of sanctions effectiveness.

Part II will reveal, however, serious methodological flaws that undermine the validity of the leading scorecards currently in use—and call into question their wide influence. Some of the problems afflicting these scorecards illustrate problems of measurement and method that arise in quantitative analyses of social phenomena generally. Others are methodological challenges that extend to all empirical studies of sanctions, including case studies. Although some of the discussion in Part II will be technical, I should emphasize that I approach the subject not as a practicing econometrician, but as an educated consumer of

¹³. In this respect I deliberately depart from Baldwin’s line of defense of sanctions as signaling and expressive devices aimed as much at domestic audiences as at the target state. David A. Baldwin, Economic Statecraft (1985) [hereinafter Economic Statecraft]. There is much, of course, to be said for sanctions as speech—mechanisms for dissociating sending states from immoral and harmful conduct by target states. My exclusion of speech values in this analysis is solely for reasons of analytical interest: sanctions almost by definition achieve their speech goals, but they do not always achieve helpful change in foreign behavior.

¹⁴. See infra Part II.A.

¹⁵. See infra Part II.B. for a detailed discussion of scorecard approaches.
quantitative analysis, addressing fellow consumers. The goal is not to try to de-bunk the use of quantitative methods in analyzing economic sanctions per se, but simply to illuminate, for the lay reader, the kinds of methodological problems that must be addressed, and solved, by any sanctions impact assessment that claims our trust as a basis for law and policy.

The discussion in Part II will focus on sanctions scorecards and, in particular, on the quantitative analysis employed by Hufbauer, Schott and Elliott in their monumental study of high foreign policy sanctions, *Economic Sanctions Reconsidered*. Those who call for "cost-benefit" analysis of sanctions may be tempted at this point to assume, or argue, that my critique is safely confined to the HSE study or, at most, to quantitative scorecards. Traditional case studies (and extrapolations from same) may continue, undisturbed by any new concerns about method.

Such complacency is misplaced. Close examination of the problems discussed below in connection with quantitative scorecards—lack of a control group, biased case selection, omitted variables, missing data, mis-specified models, multicollinearity, problems of endogeneity, problems defining the boundaries of episodes and the criteria of success, subjectivity problems, expectancy effects—will immediately reveal that *many if not all of these problems also afflict efforts to draw ordinal inferences from case studies and, by extension, efforts to predict future outcomes based on past experience*. Most of the rules of "econometrics" are, after all, simply rules of valid measurement and rules of logical inference. Measurement problems inhere in any situation (such as sanctions episodes) in which outcomes depend of the relative magnitudes of conflicting social forces ("variables"). Ignore measurement and you ignore relevant differences—whether your vehicle is a scorecard or a case study. Other "econometric" requirements—such as the requirement for including all relevant variables, specifying a plausible model, controlling for confounding factors, and taking into account feedback effects—are equally applicable to case studies. They are simply rules of logic for drawing valid inferences from the experience of history.

In short, I have chosen to focus on the HSE scorecard because it is prominent, clear, systematic and, hence, very useful in exposing the problems of method and measurement that perplex the study of sanctions generally. It is utterly wishful thinking, however, to pretend that these problems can be avoided by more anecdotes, or by continuation of the present tradition of inconsistently and partially analyzed case studies.

Most of Part II will be devoted to examining studies of sanctions themselves. However, the last section of that part will take up a different but closely related issue—the validity of extrapolating conclusions about
the effectiveness of sanctions from one area of policy to another. This part will demonstrate that the experience of "high foreign policy" sanctions (even if properly analyzed) does not apply \textit{a fortiori} to trade measures aimed at preserving the global commons. There is simply no logical or empirical basis for the USA*Engage practice of lumping environmental sanctions into the same analytical pot as "high foreign policy" measures aimed at quite different ends.

Ultimately, economic sanctions are a two-edged sword of variable sharpness. They can be potent or impotent, constructive or counter-productive, depending on the circumstances and the manner in which they are used. Even the critics of sanctions recognize this, which is why they are not proposing to outlaw economic sanctions altogether. Rather, they insist only that sanctions be confined to cases where they appear cost-effective based on prior, and continuing, analysis. The request is perfectly reasonable, in principle. What it overlooks, however, is the fact that current methods and studies do not yet provide a sound empirical or theoretical foundation for estimating the costs or effectiveness of individual sanctions either \textit{ex ante} or \textit{ex post}. In the absence of such a foundation, calls for "cost-benefit" analysis of sanctions are at best premature; at worst, a farce.

Although the focus of this article is on critiquing the current state of our "knowledge" of sanctions impact, Part III will examine briefly and preliminarily what should be done to establish a sound basis for "cost-benefit" analyses of future sanctions. It will be seen that the first and foremost requirement of sound sanctions policy is a renewed commitment to open-minded, searching, empirical examination of the actual role of economic leverage in the cases where it has been used. This examination should begin by developing a quality-controlled database of case studies involving the threat and/or use of sanctions. These case studies should be carefully structured to apply a consistent, comprehensive, and coherent analytical framework that includes the major variables that bear on sanctions effectiveness and incorporates what is known about the causal pathways by which economic sanctions influence behavior. Much more effort is needed in developing reliable measures of key variables that bear on sanctions effectiveness, and/or in

---

16. See infra Part II.C.12.
18. Id. (imposing procedural restrictions on new sanctions without proposing to prohibit existing or new sanctions measures); Kittredge Testimony, supra note 9, at 6–7 ("[W]e recognize that there will be times when sanctions will be used. We believe that we should make them as effective as possible. . . .").
finding reliable methods for assigning ordinal "scores" to subjective
variables. Precautions should be taken to guard against the well-
documented "expectancy effects" which plague all analyses involving
subjective assessments. For example, sanctions impact assessment
should not be performed (at least not exclusively) by agencies with
vested policy interests or by experts with financial ties to economic
stakeholders.

In general, my discussion will reveal that getting sanctions right will
require considerably more resources than have been devoted to sanctions
analysis to date. The standards of analysis which I call for may sound, to
some, methodologically extreme. But are they? Is it extreme to suggest
that the government that presides over a $9 trillion dollar economy—and
purports to be a leader in shaping the destiny of the world—should show
at least as much diligence in assessing a major tool of foreign policy as
the average pharmaceutical company devotes to testing the safety and
effectiveness of a hair tonic?20

I. MEASURING AND REPORTING THE COST OF SANCTIONS
to the U.S. Economy

Before examining the effectiveness of economic sanctions, it is im-
portant to review and place in perspective the evidence of their cost to
the U.S. economy. Although news stories and internet sites abound with
anecdotes of contracts and sales lost to economic sanctions, by far the
most widely cited and influential estimate of the costs of high foreign
policy sanctions to the U.S. economy overall is the one prepared by
Hufbauer, Elliott, Cyrus and Winston (HECW) in 1997.21 This section
will examine the methodology used in that study, as well as the way the
results have been reported. It will be seen that their model is plausible
and their method sound, with one small caveat. But the packaging and
reporting of their results to and by the media leaves much to be desired.
The reporting appears to have distorted and exaggerated the actual find-
ings of sanction cost.

20. 1995 GNP was $7.3 trillion. See Bureau of the Census, Econ. and Stat. Ad-
721 (1998). 1998 GNP was reported to be $8.5 trillion. See Bureau of Econ. Analysis, Gross
Domestic Product: Fourth Quarter 1998 (Final), News Release, Mar. 31, 1999, (last modi-
A. Costing Methodology and Conclusions

Rather than try to estimate the direct and indirect cost of individual sanctions episodes and then sum these costs across all episodes, HECW employ an ingenious heuristic model called the "gravity model." This model is so named for its conceptual affinity to Newton's theory which holds that the gravitational pull between two celestial bodies is positively related to the product of their masses and inversely related to their distance apart. As applied to trade, the gravity model predicts that the amount of trade between two countries, other things equal, will be positively related to the product of their outputs (analogous to size or mass) and inversely proportional to the distance between them. Since other things are not always equal, HECW insert additional variables to control for factors which might be expected to increase trade flows beyond what mere size and distance between capitals would suggest: e.g., high per capita GNP, sharing a common border, sharing a common language and/or belonging to the same trading bloc. After initially running the regressions with two-way trade as a dependent variable, HECW re-run the regressions using OECD exports and U.S. exports, respectively, as dependent variables in order to isolate the impact of foreign sanctions on U.S. businesses and workers. Finally, HECW insert nine dummy variables to reflect the existence of (1) limited, (2) moderate, or (3) extensive economic sanctions between each pair of countries which, in each case, are (1) in place in each of the measurement years 1985, 1990, or 1995, (2) ended one to two years prior to each of the measurement years, and (3) ended three to four years prior to each of the measurement years. The purpose of the lagged variables is to try to determine whether sanctions have a residual impact after they are lifted due to the so-called "unreliable supplier effect."

The authors find a highly significant correlation between sanctions and two-way trade during the years that sanctions are in place. They also find a statistically significant and negative impact on U.S. exports during the years that sanctions are in place. This is not too surprising since most such sanctions take the form of export restrictions. However, HECW find no conclusive evidence of a residual effect dampening trade in the years immediately following repeal of sanctions, though the authors acknowledge the possibility of significant residual impacts in individual sectors. In other words, residual "unreliable supplier"
effects, while they may exist in individual sectors, are not large enough to produce a statistically significant lagged decline in two-way, bilateral trade overall. 27

The estimated coefficients for sanctions provide a measure of the average effect of imposing a “limited,” “moderate,” or “extensive” sanction, respectively, in reducing U.S. exports to a target country. Applying these coefficients to actual imports in 1995 enables HECW to calculate an estimated reduction in U.S. exports to each sanctioned country, which can then be summed to yield an estimate of total exports lost in 1995 as a result of foreign policy sanctions overall. 28 These calculations yield a total estimated reduction in U.S. exports of between $15 and $19 billion in 1995. Taking Commerce Department estimates for the average labor intensity in export industries in 1992, with suitable adjustments for productivity growth between 1992 and 1995, HECW estimate that each $1 billion of lost exports would correspond to a loss of 13,800 jobs, if lost exports are not compensated by sales elsewhere. On this assumption and applying the ratio of 13,800 export jobs per billion dollars of lost exports, HECW calculate that a loss of $15 to $19 billion in lost exports would translate to 200,000 to 260,000 lost U.S. jobs. 29

In the primary report of their results, HECW are careful to point out that in a full employment economy, “lower exports do not spell an overall drop in employment.” 30 But they insist that a loss of exports does mean a loss of wages in the form of the “export sector wage premium.” Citing Commerce Department data that suggest a U.S. export wage premium of about $4,080 per worker over non-export wages, the authors compute the loss to the U.S. economy of $800 million to $1 billion per year—the $4,800 per-worker export wage premium multiplied by the putative 200,000–260,000 lost export jobs. 31

27. Id. at 6.

28. This method does not work, of course, for “extensive” sanctions which have the effect of curtailing virtually all U.S. exports to a target country. For these countries, export reductions as a result of sanctions are calculated by a variety of means: (1) for Iraq and Libya, setting the reduction at the level of U.S. exports in the last year before sanctions were imposed, adjusted for trends in world exports to the target; (2) for North Korea, setting the reduction at the U.S. share of 1995 OECD exports to South Korea multiplied by OECD exports to North Korea in 1995; (3) for Cuba, setting the reduction in U.S. exports at the U.S. share of 1995 OECD exports to the Dominican Republic multiplied by 1995 OECD exports to Cuba; and (4) for Serbia & Montenegro, setting the reduction in U.S. exports at estimated 1995 U.S. exports to all of Yugoslavia less observed exports to Croatia and Bosnia Herzegovina. Id. at tbl.8.

29. Id. at 7.

30. Id.

31. Id.
The gravity model employed by HECW has been used for nearly forty years to measure the impact of commercial policy measures on bilateral trade flows. The authors' application of it to the sanctions context appears to be basically sound. This section will discuss three main problems with the HECW sanctions cost estimates: (1) the presentation of their results to the public; (2) the authors' assumption that a loss of exports to a sanctioned country cannot be made up by exports elsewhere; and (3) the tendency of USA*Engage (though not HCEW) to apply the "cost critique" to all types of sanctions without distinction between policy contexts.

1. Presentation of results. The HECW estimate that U.S. foreign policy sanctions would cost the U.S. economy $15–20 billion per year and cost 200,000 jobs per year has been repeated in any number of advocacy statements and media reports. Seldom, if ever, do public reports of the HECW findings place them in the perspective of a $7–8 trillion U.S. gross national product in the 1995 measurement year. While the average reader might be shocked by sanctions of $20 billion per year, she might well find herself somewhat less dismayed—assuming sanctions cumulatively accomplish anything at all to enhance our national security and core values—at learning that something less than one-third of one percent of GNP has been set aside for the purpose. Moreover, the average reader's instinct would be right: the impact of sanctions on the typical American worker or consumer is much more accurately proxied by the GNP proportion than by the (seemingly large) absolute number.

Moreover, as has been seen, HECW themselves recognize that both the dollar and the job-loss figures tend to overstate the true cost of sanctions in a full employment economy. Under current conditions of full employment, significant net job loss is impossible almost by definition. In fact, as HECW acknowledge, the actual loss from sanctions in 1995 was more likely zero jobs and $800 to $1 billion per year of income—
about one-seventieth of one percent of 1995 GNP. (The income loss represents the export sector wage premium that HECW believe is sacrificed when export sanctions foreclose export opportunities.) Yet media and scholars alike have failed to disclose the lower figure which employs realistic assumptions, instead focusing exclusively on the higher figure which is based on false assumptions.  

2. The assumption of irreplaceable exports. It turns out that even the $800 million to a $1 billion cost estimate may be too high. That estimate is based on an assumption that exports denied to the target country cannot be diverted to any other country and therefore will cause layoffs in export industries, with all laid off workers taking non-export (no export premium) jobs. HECW offer no evidence, however, for the assumption that exports lost to one country cannot be diverted elsewhere. Surely at least some can, and to that extent the export wage premium is conserved. The export wage premium is also conserved to the extent that workers displaced from one export company take a new job in another export-oriented company.

As against this upward bias in reported cost estimates, there is the downward bias associated with the fabled “unreliable supplier effect”: the tendency of foreign purchasers to abstain from placing new orders with companies deemed to be “unreliable suppliers” due to the export policies of their parent government. Failure to account for this effect would tend to under-state actual costs of sanctions. How significant is the unreliable supplier syndrome? The phenomenon so far has proved to be as elusive as the so-called “competitiveness” harm caused by high domestic environmental regulation. Companies swear by it; economists can’t find it in the numbers. Under these circumstances, the most that
can be said, reliably, about the cost of foreign policy export sanctions to the U.S. economy is what can be gleaned from anecdotes: certain companies probably lost some attractive sales and/or investment income that might have been earned through dealings with a few highly problematic countries. This state of affairs may raise issues of equity and even claims for compensation of those companies. The fact remains: the leading estimate currently available offers no support for the proposition that the cost of export sanctions to the U.S. economy is large enough to be a matter of national concern even in the high foreign policy realm where sanctions are both large and frequent.

3. Extrapolation from high foreign policy to environmental sanctions. Even if one concludes that wide-ranging “foreign policy” sanctions are costly to U.S. businesses and workers, this does not prove that environmental sanctions—a term I use to refer to sanctions aimed at conserving the global commons—are also costly. The latter are relatively rarely applied and, when applied, take a quite different form: they are narrowly targeted on particular products and they restrict U.S. imports, not exports. So far, the evidence is clear that environmental trade embargoes have had no deleterious effect on U.S. consumers or workers. Based on the record to date, USA*Engage is simply off base to

---


39. For anecdotal accounts of lost sales see High Cost of Unilateral Sanctions, supra note 37, at 3–6; Fites, supra note 36; and President’s Export Council Review, supra note 38, at app. II. These anecdotes, though cumulatively too lengthy to recount here, should not be discounted. They strongly suggest lost sales. Moreover, many of these anecdotes involve failed bids and lost business opportunities due to the threat of future sanctions that never actually materialize from imposed sanctions. Such losses are not captured by the HECW lagged variable estimate of residual trade impacts. To that extent, the HECW estimate may tend to understate the true cost of sanctions—though the exact magnitude of the bias is hard to pin down due to the difficulty of determining in each case whether the contract really would have gone to the U.S. supplier but for threat of sanctions.

40. While foreign policy sanctions have been employed on over 200 occasions, often covering a very wide range of goods, environmental sanctions have been applied in only three cases and, in each case, surgically: (1) on imports of purse-seine-caught yellowfin tuna harvested in the Eastern Tropical Pacific (to protect dolphins); (2) on turtle-associated shrimp caught without turtle excluder devices (to protect endangered sea turtles); and (3) very briefly, on selected wildlife products from Taiwan (to secure Taiwan’s cooperation in protecting endangered rhinos and tigers). Environmental sanctions have been explicitly or implicitly threatened rather more often. See Hudec, supra note 3, at 98–106. The tuna-dolphin import
complain of the high and immediate economic cost of "sanctions" to U.S. businesses and workers without making any distinction between sweeping national security export sanctions, on the one hand, and narrowly-focused import restrictions imposed for environmental purposes, on the other.41

II. MEASURING THE EFFECTIVENESS OF TRADE SANCTIONS

Having briefly reviewed current estimates of the cost of economic sanctions, I now turn to the controversy over their effectiveness. A moment’s reflection will reveal that there are three main ways one might go about empirically substantiating assertions about the effectiveness of trade sanctions in changing target state behavior: anecdotes, case studies,

restrictions had a significant economic impact on foreign tuna fishers and processors but did not materially affect U.S. canned tuna prices or jobs. See Richard W. Parker, The Use and Abuse of Trade Leverage to Protect the Global Commons: What We Can Learn from the Tuna-Dolphin Conflict, 12 GEO. INT’L. ENVTL. L. REV. 1, 34–36 (1999) [hereinafter, Use and Abuse of Trade Leverage]. The shrimp-turtle embargoes applied only to Malaysia, Thailand, India and Pakistan (19 other nations avoided sanctions by complying with U.S. conservation standards) and appear to have had no impact on U.S. prices or jobs. See Report of the Panel, United States—Import Prohibition of Certain Shrimp and Shrimp Products, WT/DSS58/R, at ¶ 1.1 and 2.16 (May 15, 1998) [hereinafter, Shrimp-Turtle Panel Report] (review of countries embargoed and certified); telephone interview with B.G. Thompson, Fisheries Consultant (Dec. 29, 1999) (explaining that the embargoes had no price impact on consumers because “75% of shrimp marketed from the U.S. is grown through aquaculture. This is the shrimp that establishes the market price of shrimp”; and because most imported shrimp was certified by the State Department as meeting U.S. standards.) If anything, the embargoes benefited U.S. producers by applying the same standards to foreign producers that already applied domestically in the U.S. In the rhino-tiger case, the United States imposed about US $25 million worth of sanctions on the import of Taiwanese fish, leather and similar goods (against a Taiwanese annual foreign trade of US $162 billion) in 1994, and repealed the sanctions in 1995 when Taiwan agreed to strengthen enforcement of its illegal trade laws. See Simon Beck, U.S. Recognises Moves to End Wildlife Trade, SOUTH CHINA MORNING POST, Sept. 13, 1996, at 4; Amy E. Vulpio, Note, From the Forests of Asia to The Pharmacies of New York City: Searching for a Safe Haven for Rhinos and Tigers, 11 GEO. INT’L. L. REV. 463, 480 n.138 (1999). In short, no harm to U.S. consumers or workers has been identified in relation to any of these environmental trade measures to date.

41. Hufbauer et al. are quite careful to limit their observations and recommendations to the “high foreign policy” context whence their data is drawn. Others are not so careful, however. The National Association of Manufacturers, for example, simply asserts that the use of “trade sanctions for environmental protection purposes . . . is a new [U.S.] national security objective,” CATALOG OF NEW SANCTIONS, supra note 4, at 2, thereby ignoring a host of important distinctions between the two types of sanctions relating to scope, method of operation, cost, political dynamic, and effectiveness. See discussion infra Part II.C.12. Likewise, the President’s Export Council uncritically lists sanctions linked to environmental activity alongside measures aimed at countries that harbor war criminals, abet terrorism, or fail to cooperate on narcotics enforcement. See PRESIDENT’S EXPORT COUNCIL REVIEW, supra note 38, at app. I, II.A (terrorism), II.B. (narcotics), II.C (environmental activities), II.D (harboring war criminals).
and what I will call "scorecards." Each of these methods, it will be seen, has its own strengths and weaknesses. After briefly outlining the rather obvious drawbacks of anecdotes and case studies as tools of empirical analysis, this section will provide a detailed critique of the methodology used in current scorecards of sanctions effectiveness. While some of the problems so identified may be resolvable by more sophisticated methods, this section will conclude that any serious effort to arrive at reliable measures of effectiveness will need to start by strengthening both our theoretical understanding of how sanctions work, and the database of case studies on which all scorecards necessarily draw.

A. Stories and Studies

Anecdotes of sanction failures—of contracts lost, of human suffering, and of policy failure—are popular with journalists, speech writers, and scholars alike for obvious reasons. They make good leads for books and articles. They illustrate points vividly and concisely but not too laboriously. Accumulation of several anecdotes in support of a point can give the impression of broad-based empirical support: one is a prime number, three is a prime number, so are five and seven, ergo, all odd numbers are prime. Yet, while anecdotes are highly valuable

---

42. Again, I use the term "effectiveness" to mean the contribution of sanctions to achieving desired changes in target state behavior—excluding any symbolic, expressive, or domestic political function—for the reasons set forth in note 13 supra and accompanying text.

43. The proclivity of journalists and editorialists for using anecdotes to support their claims will be obvious to any newspaper reader. See, e.g., Snake Oil, supra note 6; Drezner, supra note 33; Louis Uchitelle, Who's Punishing Whom?: Trade Bans are Boomerangs, U.S. Companies Say, N.Y. TIMES, Sept. 11, 1996, at D1, for only a few of the myriad possible examples of anecdotalism in policy analysis. Of course, the format of most news and magazine stories inherently precludes detailed treatment of either individual cases or the history of sanctions generally. "Anecdotalism," as I use the term, is defined not by the fact that historical episodes are treated briefly in short-format pieces, but by the tendency of some authors and essayists to draw their own inferences and conclusions from the anecdotes they cite, rather than ground their conclusions on close, contextual and nuanced analysis of in-depth case studies done by themselves or others. See, e.g., Richard N. Haass, Economic Sanctions: Too Much of a Bad Thing, (visited Feb. 3, 2000) <www.usaengage.org/studies/haass1.html> [hereinafter, Too Much of a Bad Thing]; Robert A. Pape, Why Economic Sanctions Do Not Work, 22 INT’L SEC., Fall 1997, at 90. Anecdotalism defined in this manner is neither confined to, nor excused by the confines of, short-format writing. For a prominent example of anecdotalism employed in the book format, see ABRAM CHAYES & ANTONIA CHAYES, THE NEW SOVEREIGNTY: COMPLIANCE WITH INTERNATIONAL REGULATORY AGREEMENTS (1995) (relying heavily on their own anecdotal accounts of trade sanctions episodes to establish the inefficacy of trade sanctions).

44. This is the (obviously intended) effect of the President’s Export Council Review, which dedicates an Appendix entirely to the cataloguing of lost sales from unilateral economic sanctions—even though the anecdotes appear to be stories culled uncritically from a handful of disappointed corporate bidders on foreign contracts (United Technologies, Boeing,
as illustration, they have only weak probative value for reasons that are equally obvious: anecdotes are, by definition, isolated examples deliberately (not randomly) picked from the database of history, stripped of context and nuance by virtue of their brevity and "spun" into a form and substance suitable to proving a point. Thus, the failure of sanctions to topple Castro is often cited as support for the inefficacy of economic sanctions employed for high foreign policy purposes.\(^ {45} \)

Yet the alert citizen and policy-maker might well ask, did sanctions nonetheless weaken Castro's ability to finance counter-revolutionary movements in Latin America; did they force the Soviet Union to deplete its own resources subsidizing Castro; and/or did economic sanctions force Castro to liberalize his state-run economy more than he otherwise might have done?\(^ {46} \)

Did the hardships imposed by sanctions on Cuba serve to deter other countries from choosing the "Communist" and expansionist path during those years when international communism still had some charisma overseas? Such questions are hard to answer satisfactorily in the sound-bite format demanded by anecdotal approaches.

Case studies, on the other hand, restore much of the detail, nuance, and context omitted by anecdotes. They are an indispensable foundation for any and all empirical and theoretical analysis of sanctions.\(^ {47} \) But case studies have their own obvious deficiencies. They tend to bog the reader down in a mass of detail. They yield (quite often) highly nuanced conclusions with no sound-bite potential whatever. Case studies offer, at the end of what may be book-length analysis, only the "lessons" of a single case.\(^ {48} \) And no effort appears to have been made, to date, to develop a

---

45. See, e.g., Too Much of a Bad Thing, supra note 43, at 5.

46. For an affirmative answer to these subtler questions see Susan Kaufmann Purcell, Cuba, in Economic Sanctions and American Diplomacy 35, 35-57 (Richard N. Haass ed., 1998). The alert reader may also think at this point to ask, is the anecdote about Cuban sanctions indicative of the problem of anecdotal approaches generally, or is the author cherry-picking from the history of the use of anecdotes? But that quandary just further corroborates my point about the deficiencies of the genre.

47. See discussion infra Part III.

standard analytical format for case studies. The lack of empirical standardization makes it very difficult to make inter-case comparisons and contrasts, to account for confounding variables or, ultimately, to extract from case studies general insights into the factors that contribute to or undermine sanctions effectiveness.

The deficiencies of anecdotes and case studies as empirical tools lead straight to the attraction of scorecards.

B. Scorecards

The defining feature of "scorecards," as I use the term, is that they rate the success of sanctions across a wide spectrum of cases in a given issue area. Besides tabulating success or failure rates, scorecards also may examine the correlation between certain variables which are thought to contribute to success or failure, on one hand, and the actual success or failure of sanctions, on the other.49 Most of all, scorecards employ quantitative methods to test hypotheses and report results. Their methods and conclusions are subject to validation by seemingly objective criteria: the standards of sound statistical analysis.50

Scorecards are attractive, then, precisely because they appear to solve many of the problems which bedevil anecdotal accounts and case studies. By examining all (or all known) cases in a particular issue area, they avoid the problems of cherry-picking which plague anecdotal approaches. By scoring cases across a range of variables on the basis of detailed case studies, scorecards can claim, at least, to be able to detect general patterns of behavior that emerge from the welter of detail that fills case studies. Most of all, scorecards are appealing because they distill the experience of a vast range of cases into a few easy-to-report, easy-to-remember variables and associated numbers which seem to carry the authority of "science."51 If there is any explanation for the influence


49. This is, of course, a primary aim of the HSE study. See ECONOMIC SANCTIONS RE- CONSIDERED (1st ed.), supra note 1.

50. For lucid and readable introductions to statistics and econometric methods see, respectively, DAVID FREEDMAN ET AL., STATISTICS (3d ed. 1998) [hereinafter STATISTICS], and WILLIAM E. GRIFFITHS ET AL., LEARNING AND PRACTICING ECONOMETRICS (1993) [hereinafter PRACTICING ECONOMETRICS].

of scorecard approaches on scholarship and public policy debate, surely that explanation lies in the brevity, comprehensiveness, and seeming methodological rigor that scorecards seem to afford.

Trade sanctions scorecards have been compiled in at least three major subject areas: (a) high foreign policy; (b) trade policy; and (c) environmental protection. Before critiquing the leading scorecards it may be helpful to describe them briefly.

High foreign policy sanctions. By far the most influential empirical study of the effectiveness of economic sanctions is Hufbauer, Schott and Elliott’s (HSE’s) monumental work, Economic Sanctions Reconsidered.52 The work, and/or its lead author, Gary Hufbauer, has been cited in at least 85 national magazine stories, 100 major newspaper articles or editorials, and 66 law review articles since it first appeared.53

In the first edition, HSE employ econometric regression techniques to examine 103 cases between 1914 and 1985 in which economic sanctions have been applied for five sorts of foreign policy purposes:

1. Destabilizing foreign governments;
2. Disrupting military adventures;
3. Impairing the military potential of a hostile state;

ability to displace fallible human judgement, and their role in imposing “commensurability” on otherwise disparate values, thereby facilitating “rational” choice.


The Problem with Scorecards

(4) Achieving "major" policy change; or
(5) Accomplishing "modest" policy changes.54

The economic sanctions at issue may assume a variety of forms: import and/or export embargoes, withdrawal of economic or military assistance, or financial asset freezes.55

HSE measure success in each case by first comparing the sender's goal with the actual outcome, to yield a policy result score scaled from 1 (complete failure) to 4 (complete success).56 They then assign a sanctions contribution score also scaled from 1 (sanctions irrelevant or counterproductive) to 4 (sanctions a major contributor to success).57 Multiplying the two scores together yields an overall "success score" scaled from 1 (outright failure) to 16 (outright success), with a score of 9 or higher rated a "success."58 Applying this scoring method to the evaluation of case studies of 103 cases over the period 1914-1985, HSE come up with an overall success rate for sanctions: 36 percent success overall, and 40 percent success rate for sanctions involving "modest policy goals."59

The conclusion of Economic Sanctions Reconsidered—that sanctions actually "worked" in over a third of the cases—was considered revisionist at the time (the consensus in scholarly circles then being that economic sanctions almost never work).60 This revisionist finding then led HSE to enlarge the focus of their analysis beyond the threshold question, whether sanctions ever work, to a more fruitful inquiry into why sanctions work or fail: what situations and methods of application are most favorable, and least favorable, to success.61 That, of course, is precisely the inquiry that USA*Engage is now calling for as a precondition of new sanctions.62

55. See id. at 36-37. The authors define "economic sanctions" as "the deliberate, government-inspired withdrawal, or threat of withdrawal, of customary trade or financial trade relations." Id. at 2. However, as seen infra note 89 and accompanying text, their database includes only six threat-only cases, an impossibly low number in the context of 115 actual, imposed sanctions. So it would probably be more accurate to delete the phrase "or threat of withdrawal" from their definition.
57. Id.
58. Id.
59. Id. at 79-80.
60. As David Baldwin observed in 1985, one "salient characteristic of the literature on economic statecraft is the tendency to denigrate the utility of such techniques." Economic Statecraft, supra note 13, at 55. Baldwin supports this assertion with a long list of quotations and cites to authors denying the effectiveness of economic sanctions. Id. at 55-59.
61. See discussion accompanying note 66 infra for their conclusions on the conditions favoring and disfavoring successful sanctions.
62. See Kittredge Testimony, supra note 9, at 5-7.
HSE pursue this broader inquiry by employing econometric techniques to test various hypotheses about the factors ("variables") which might be thought to make sanctions more or less effective. That is, they correlate the dependent variable (success) on the left-side of the equation with at least 15 independent variables on the right hand side, and look at the coefficients and standard errors that result.63

Besides relying on multivariate regression analysis for insights, HSE employ a technique known as "cross-tabulation," which involves comparing the average score for each independent variable in successful and failed cases, respectively.64 For example, the authors report that sanctions have been successful in 52 percent of cases where the policy goal is to de-stabilize a foreign government, and in 33 percent of cases where the policy goal is disruption of military adventures or modest policy change. From this the authors conclude that sanctions deployed for de-stabilization purposes are "surprisingly successful." However, they also note the "average index of economic health and political stability for target governments" is 1.4 in failed cases (on a scale of 1 to 3) and 1.9 in successful cases. From this the authors conclude that governments in distress are more easily de-stabilized. Indeed, the second (1990) edition employs cross-tabs analysis exclusively.65

On the basis of such analysis the authors adduce "Nine Commandments" which they suggest designers of sanctions should comply with if they hope to be successful. The Nine Commandments, essentially unchanged in the 1985 and 1990 editions, are:

(1) Sender governments should not ask too much of sanctions (they are seldom effective in achieving major policy change or impairing the military potential of a major power).
(2) Sanctions requiring international cooperation are seldom successful, while those seeking modest policy changes often do not require cooperation from allies to succeed.

(3) Sanctions against politically or economically unstable countries are more likely to succeed than those targeted at healthy, stable countries.

(4) Economic sanctions aimed at friends and allies are more effective than those directed against long-time adversaries of the sender country.

(5) The longer the duration of sanctions, the lower the likelihood of success.

(6) The more costly the sanctions to the target state, the greater the likelihood of success.

(7) The more sanctions cost the sender, the less likely they are to succeed.

(8) Companion measures—covert action, quasi-military or regular military operations—are used most frequently in conjunction with sanctions involving destabilization or impairment of military potential. Their effect may swamp the impact of sanctions (causing sanctions to be rated as a "failed" contribution even when sanctions may have augmented the impact of companion policies).

(9) Sender governments should think through their means and objectives carefully before taking a final decision to deploy sanctions.66

Though the HSE study initially made waves for reporting that sanctions work more frequently than had been expected, the authors’ own appraisal of the instrument has diminished over time. Indeed, even in the first edition (1985) the authors had noted that foreign policy sanctions seemed to be losing effectiveness: while nearly half of all sanctions begun before 1973 had succeeded, only 25 percent of those begun after 1973 were rated successful.67 From this, HSE concluded that while "sanctions occasionally bear fruit . . . when planted in the right soil and nurtured in the proper way" they are a "decreasingly useful policy instrument."68


68. *Id.* at 81.
more emphasis in the second edition.\textsuperscript{69} Perhaps the spectacular failure of Iraqi sanctions—after Hufbauer had publicly predicted success—further eroded his confidence in economic statecraft.\textsuperscript{70} In any case, by 1998 Hufbauer was calling economic sanctions the "snake oil of diplomacy" and he himself had hired on as a consultant to the USA*Engage sanctions reform campaign.\textsuperscript{71}

\textit{Trade policy scorecards.} The success of \textit{Sanctions Reconsidered}—and the intrinsic appeal of the method it employs—has inspired at least three additional sanctions scorecards. Two of these, one by Alan Sykes and another by Thomas Bayard and Kimberly Elliott (BE), examine the use of U.S. trade leverage to open foreign markets under Section 301 of the Trade Act of 1974.\textsuperscript{72} In examining 83 trade policy disputes over the period 1974–1990 both studies adopt an essentially binary definition of "success." Unlike the HSE study, neither of the trade policy studies attempts to assign values to the separate \textit{contribution} of sanctions to success.\textsuperscript{73} Rather, these studies simply examine correlation between certain independent variables, on one hand, and success or failure, on the other.\textsuperscript{74} The main distinction between the studies, beyond the fact that

\begin{itemize}
  \item \textsuperscript{69} See \textit{Economic Sanctions Reconsidered} (2d ed.), \textit{supra} note 1, at 105–14.
  \item \textsuperscript{70} See discussion \textit{infra} notes 188–190 and accompanying text.
  \item \textsuperscript{71} \textit{Snake Oil}, \textit{supra} note 6.
  \item \textsuperscript{72} \textit{Aggressive Unilateralism}, \textit{supra} note 2; \textit{Constructive Unilateral Threats}, \textit{supra} note 2.
  \item \textsuperscript{73} Bayard and Elliott claim that they employ a somewhat stricter definition of success than Sykes, which partly accounts for their lower success rate. \textit{Aggressive Unilateralism}, \textit{supra} note 2, at 695.
  \item \textsuperscript{74} Sykes considers the impact of six factors: (1) whether or not the cases involves an alleged GATT violation; (2) whether or not a panel was convened and whether it ruled in favor of the United States; (3) whether the case involved export promotion; (4) whether the case involved agriculture (an area deemed particularly problematic); (5) whether or not the recipient was a developing country dependent on the United States for special trade preferences (a circumstance thought to heighten U.S. bargaining power); and (6) whether or not the U.S. actually retaliated in the case. \textit{Constructive Unilateral Threats}, \textit{supra} note 2, at 318, app.

BE estimate the impact of five quite different factors derived from a game-theoretic model devised by John McMillan: (1) economic stakes for the sending state; (2) likelihood of counter-retaliation; (3) proportion of target's total exports sent to the United States; (4) the nature of the measure complained of (border restriction, subsidy, or technical/administrative barrier); and (5) the credibility of the threat. The first variable, economic stakes, is proxied by a dummy which takes the value 1 if the U.S. pre-sanction trade pattern is between $10 million and $100 million, and zero otherwise, on the theory that small cases (involving products with below $10 million of pre-sanctions trade) are less credible predicates for retaliation because the U.S. government is simply less concerned with them, while very large pre-sanctions trade flows are also less credible situations for retaliations because in those cases the U.S. will fear counter-retaliation against U.S. exports. \textit{Aggressive Unilateralism}, \textit{supra} note 2, at 698–700, tbl. 4. Factor (2)—likelihood of retaliation—is proxied by a dummy variable, ECCAN, which takes the value 1 if the target is EC or Canada, 0 otherwise, on the theory that these are the only two countries that have ever counter-retaliated against a U.S. trade retaliation in the past. The indices for factors (3) and (4) are self-evident. Factor (5)—credibility of threat—is
they focus on quite different variables, lies in their basic method of inference. Sykes employs simple “cross-tabulation” techniques to derive rough, heuristic inferences from the data—such as the interesting insight that cases involving alleged GATT violations do not have a significantly higher success rate than cases in which the U.S. acts completely outside the GATT framework. BE, by contrast, attempt a full-fledged “probit” regression analysis to try to determine the relative importance of the independent variables in determining successful outcomes.

Sykes finds “success” in 31 of 35 completed cases in which the U.S. alleged a breach of the GATT, and in 27 of 33 completed cases in which the U.S. complained of an “unreasonable” practice not involving an alleged violation of the GATT. BE, using a somewhat stricter criterion of success, conclude that “Section 301 has led to at least partial market opening in just over half the cases overall, and in two-thirds of the cases ending after the announcement of the President’s trade policy action plan in September 1985 [a “get-tough” plan].” The rather optimistic results obtained by studies of the efficacy of trade policy sanctions may account in some part for the decision of the USA*Engage coalition to exempt trade policy sanctions from their critique and “reform” proposal.

The Charnovitz environmental policy scorecard. Steve Charnovitz has compiled a very informal scorecard of the track record of environmental trade leverage (ETL) in protecting the global commons under the authority of the so-called Pelly Amendment. Charnovitz finds a “58 percent success rate” despite the fact that sanctions were never actually imposed (only threatened) in the cases comprising his database. The Charnovitz scorecard is a very useful first cut at the issue, but its empirical pretensions are few. It makes no effort to ascertain the factors proxied by two dummy variables which take the value 1 if the sanctions episode occurred after the September 1985 Trade Policy Action Plan, or after the 1988 Competitiveness Act, respectively—the idea being that both of these events were “get-tough” signals of U.S. resolve to foreign governments). Id.

76. Aggressive Unilateralism, supra note 2, at 691–92, 695–96.
77. Constructive Unilateral Threats, supra note 2, at 310, 314.
78. Aggressive Unilateralism, supra note 2, at 687.
80. Steve Charnovitz, Recent Developments: Environmental Trade Sanctions and the GATT: An Analysis of the Pelly Amendment on Foreign Environmental Practices 9 Am. U. Int’l L. & Pol’y 751 (1994) [hereinafter Analysis of the Pelly Amendment](examining 14 episodes over the period 1973–1993 involving U.S. invocation of the Pelly Amendment to the Fisherman’s Protective Act to threaten trade sanctions against a foreign country for “undermining an international conservation agreement.”)
81. Id. at 773.
bearing on the success of ETL, nor does it employ any particular technique more sophisticated than scoring and counting to tabulate results. Its case coverage is not comprehensive and, in fact, omits the two most important trade and environment cases of the era. But it does suggest that U.S. sanctions or threats deployed to protect the global commons have achieved at least some of their objectives on a number of occasions.

C. Critique of the Scorecard Methodology

All of the scorecards reviewed above represent valuable and important contributions to the literature on sanctions. The impressive enterprise of these authors in collecting an enormously wide range of cases, analyzing the primary studies, coding variables, and tabulating the results reflects a commitment to systematic exploration that should, and no doubt will, serve as both a starting point and inspiration for all future scholarship in this important field. However, the results of these studies have been widely reported without, so far, sufficiently rigorous evaluation of the methods used to produce those results. As a result, the

82. These are the tuna-dolphin and shrimp-turtle disputes, both landmark disputes that produced WTO litigation over the U.S. use of trade sanctions to protect global commons resources. See Hudec, GATT Legal Restraints, supra note 3, at 98–106.

83. See Analysis of the Pelly Amendment, supra note 80, at 806 (finding 12 of 18 cases of threatened environmental sanctions at least partly successful in achieving conservation goals).


The international law and foreign policy community appears to have largely accepted the HSE findings without major reservation. See, e.g., Kenneth W. Abbott, Coercion and Communication: frameworks for evaluation of economic sanctions, 19 N.Y.U. INT’L L. & POL. 781, 787 (1987) [hereinafter, Coercion and Communication](observing that “Hufbauer and
“numbers” derived from these studies have been allowed to sway scholarly and public opinion unguarded by the necessary caveats as to their reliability. Focusing on the widely-disseminated HSE and BE scorecards, this section will reveal serious problems of measurement and method which need to be borne in mind when evaluating their findings and recommendations.

The specific methodological problems to be reviewed here fall into twelve main categories: (1) lack of control group; (2) selection bias; (3) subjectivity; (4) proxy problems; (5) problems of defining the boundaries of “episodes” and “success”; (6) model mis-specification; (7) problems of endogeneity; (8) collinearity; (9) the difficulty of modeling dynamic processes by static proxies; (10) invalid cross-tabs comparisons; (11) insufficient data and/or unreliable data; and (12) invalid cross-contextual extrapolation.

1. Lack of Control Group or Baseline

The scorecards mentioned above are generally understood to be quantitative estimates of the “effectiveness” of trade leverage in achieving foreign policy, trade policy, or environmental policy goals. In fact, none of them contain actual estimates of the contribution of trade sanctions per se to policy outcomes. To achieve a true estimate of the contribution of sanctions to “successful” outcomes, one would have to compare success rates in cases where trade leverage is used with success rates in a control groups comprised of cases where leverage is not used.85 The above scorecards contain no such control group, but rather examine only cases where trade leverage was used. As a result, these scorecards should be understood not as measures of the effectiveness of sanctions per se in contributing to favorable outcomes, but rather as measures of the contribution of various subordinate variables (such as multilateral cooperation with sender, size of sender, etc.) to the effectiveness of sanctions in the cases where sanctions are used.86

Schott’s methodology is a well-accepted way to produce precise analyses of complex and variable data” but that it “requires a great deal of subjective judgment ... even if one accepts the author’s analytical framework, other researchers could reach very different conclusions.”); Leopoldo Lovelace, Economic Sanctions Reconsidered (2d ed.), 87 AM. J. INT’L L. 178 (1992) (book review offering no criticism of methodology); William Diebold, Jr., Economic Sanctions Reconsidered, 64 FOR. AFF. 365 (Winter, 1985/86) (book review opining that “this study is a good antidote to the loose generalizations common to this subject”); Noland, supra note 11, at 86–88. As will be seen infra, problems of subjectivity are only the tip of the methodological iceberg with which HSE have collided.

85. Indeed, strictly speaking, this would prove only association, not causation or contribution. See STATISTICS, supra note 50, at 150.

86. Logically, one might conceivably arrive at a zero-threat estimate of sanctions contribution by extrapolating from a regression line which correlates size of threat with favorability
A further consequence of failing to include a control group of sanction-free cases is to sweep under the rug a key question urgently put forward by defenders of sanctions: what are the alternatives to sanctions and how “successful” are they?\footnote{See ECONOMIC STATECRAFT, supra note 13, at 123 (1985) ("The most common and most serious shortcoming in most assessments of the utility of techniques of statecraft is failure to cast the analysis in comparative terms.").} Suppose it is true, for example, that foreign policy sanctions are successful, by some measure, only about 20 percent of the time. Is that the same success rate as would occur without sanctions, twice the rate, four times the rate, or half the rate? We simply cannot know from these scorecards, because of the way they are designed. As a result, we have no basis to draw valid inferences about what the success rate of diplomacy might be in a world without economic sanctions.

2. Selection Bias

One of the foundational premises of logical inference based on sampling a portion of a larger population (here, the population of all cases where economic leverage is applied) is that the sample is random. As Siegelman and Donohue have pointed out, a researcher who studies a non-random subset of the universe of cases is “necessarily studying both the selection mechanism and the underlying population of cases simultaneously... As an eminent statistician once wrote, ‘If you catch fish with a net having a 6-inch mesh, you are liable to formulate the hypothesis that all fish are more than 6 inches in length.’”\footnote{Peter Siegelman & John J. Donohue III, The Selection of Employment Discrimination Disputes for Litigation: Using Business Cycle Effects To Test The Priess-Klein Hypothesis, 24 J. LEGAL STUD. 427, 429 (1995) (quoting Irving John Good, Fallacies, Statistical, 5 INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES 292 (1968)).}

Here, the scorecards under review are constructed to include only cases which involve either the actual application of sanctions (in the case of HSE, with six exceptions),\footnote{It is hardly conceivable that this tiny sample encompasses the universe of threat-only cases in the foreign policy realm over that period. The six threat-only cases are identified as such by a “dash” notation in the “cost-to-target” column in Tables 4.3–4.7 of ECONOMIC SANCTIONS RECONSIDERED (2d ed.), supra note 1, at 84–90. The cases are: USSR v. Romania (62–3); U.S., Canada v. South Korea (75–1); League of Nations v. Yugoslavia (21–1); League v. Greece (25–1); Western Allies v. GDR (61–3); and U.S. v. El Salvador (87–3). This tiny sample excludes even some of the more famous and overt threat cases, such as U.S. linkage of China’s Most Favored Nation (MFN) trading status to improvement of China’s human rights record.} or the issuance of formal, publicized
The Problem with Scorecards

Largely excluded from the net are cases where the sender quietly backs down or simply decides not to issue a threat or open a public investigation, but nonetheless makes it clear to the target that it finds the filing meritorious. Also excluded are cases where the target quietly accedes to the sender’s minimal demands, before full formal adversarial processes commence. The former cases, I would suggest, are minor failures: no public confrontation results, no trade is restricted. The latter category excludes, by definition, all the greatest successes of sanctions: cases where the sender’s goals were achieved without trade restriction or public confrontation. The practical reasons for excluding quiet threat cases are clear and understandable: such cases are hard to spot, and finding them would require an enormous

90. The Sykes/BE studies of trade policy sanctions include only those episodes in which a section 301 case was filed and investigated or self-initiated by USTR, thereby excluding most, if not all, cases involving only behind-the-scenes or tacit threats. Constructive Unilateral Threats, supra note 2; Aggressive Unilateralism, supra note 2. Charnovitz likewise chooses to examine only cases in which the U.S. formally “certified” another country, under the Pelly Amendment, as undermining an international conservation agreement. Certification is, of course, a very formal and public event. As Charnovitz observes, “a number of countries took action following a threat of Pelly certification, and thus were never certified. These ‘successes’ are not tallied here.” Analysis of the Pelly Amendment, supra note 80, at 141-42.

91. Indeed, threat cases are the “normal” sanctions scenario except where one or more states miscalculate. As Eaton and Engers have observed, “In a world of perfect information, senders never actually resort to punishing. If the measure works, the target shapes up in anticipation. If not, the sender never threatens it initially. Hence, the very imposition of measures means that something went awry: either the sender underestimated the target’s cost of compliance (and sanctions eventually fail), or the target underestimated the sender’s resolve (and sanctions ultimately succeed).” See Jonathan Eaton & Maxim Enger, Sanctions: Some Simple Analytics, Papers and Proceedings of the 111th Meeting of the American Economics Association, New York, New York, Jan. 3–5, 1999, in 89 AM. ECON. REV. 409–10 (1999).

Two examples of successful quiet leverage may illustrate the point. The author, while serving as Assistant General Counsel at USTR, led a U.S. delegation in negotiating a trade agreement, which opened the Japanese market to over $400 million in exports of U.S. surimi (a high-value processed fish product). That lucrative deal was negotiated without any public threats, but with the aid of quite pointed and explicit private threats that both GATT litigation and trade sanctions under section 301 would follow if the dispute was not resolved. This “quiet episode” does not show up in any trade sanctions scorecard, but it stands as an unequivocal success story for the use of trade leverage.

In the realm of environmental policy, the United States obtained the cooperation of Panama, Honduras and Belize with the Inter-American Tropical Tuna Commission’s dolphin conservation program without formal Pelly certification, by simply writing a letter to these governments, observing that their vessels were fishing in violation of the program and reminding the governments of U.S. policy of imposing trade sanctions on countries which allow their vessels to undermine international fisheries conservation agreements. Likewise, the mere threat of a Pelly certification was sufficient to persuade Ecuador to adopt a strict dolphin-safe policy in order not to imperil its large shrimp trade with the United States. See Use and Abuse of Trade Leverage, supra note 40, at 34, 53. Neither of these “quiet-threat” episodes, however, qualify for any existing scorecard.
amount of rummaging through diplomatic history. The fact remains: a very important subset of cases has been effectively excluded. The net effect is to bias downward the measured success rate for sanctions, particularly for those sanctions with modest aims (trade policy, environmental policy, and modest foreign policy goals) that are most susceptible to quiet diplomacy.

3. Subjectivity Problems

By far the most frequent criticism of the HSE scorecard and, indeed, scorecards generally, is the subjectivity often involved in “scoring” dependent and independent variables (i.e. sanction effectiveness and the putative causes thereof). HSE try to measure the “effectiveness” of sanctions by rating outcomes on a scale of 1 to 4, and by rating contribution of sanctions to outcomes, also on a scale of 1 to 4. The resulting scores for degree of success and degree of contribution are then multiplied to yield a score between 1 and 16, with overall scores greater than 9 rated a success. HSE also employ subjective scoring on the other side of the equation to reflect factors thought to determine effectiveness: warmth of prior relations between sender and target (scored 1 to 3), economic health and stability of target (scored 1 to 3), cost to sender (scored 1 to 4), international cooperation with sender (scored 1 to 4).

Where do these numbers come from? The scoring of such variables requires highly subjective judgments on at least three levels: first, the decision as to which case studies to report in the database; second, the decision as to which passages from each case study to excerpt in the two-or-three-page write-up of each case; and, third, the decision on the “score” chosen to encapsulate such complex concepts as sanction success, contribution to success, health and stability of the target, warmth of prior relations, etc. Nowhere is the basis for any scoring disclosed to the reader. Indeed, in some cases, the authors’ very scanty excerpts reveal contradictory assessments of sanction’s success and/or contribution.

92. See Coercion and Communication, supra note 84, at 787; When Sanctions Make Sense, ECONOMIST, Aug. 3, 1985, at 59 (“the scoring [in Economic Sanctions Reconsidered] is vague and multiplying the marks compounds any errors”; Olson, supra note 84, at 324 (“brief discussion of . . . [the basis for HSE’s subjective success scoring] leaves this reviewer uncomfortable”).

93. ECONOMIC SANCTIONS RECONSIDERED (1st ed.), supra note 1, at 32-33.

94. See discussion supra notes 63 and 74.

95. Take case 74-1, which I chose at random by simply opening the HSE study and examining the first case I saw. ECONOMIC SANCTIONS RECONSIDERED (1st ed.), supra note 1, at 491-95. It involves U.S. sanctions against Turkey in the wake of the Turkish invasion of Cyprus, 1974-78. HSE score the episode a failure for outcome (score of 1) and a failure for contribution of sanctions (score of 1). This scoring is based on two case studies—one by Keith R. Legg and another by Ali S. Karaosmanoglu—which, if anything, seem to support
While laymen might assume that quantitative methods are being used to determine the effectiveness of sanctions, the latter is not being measured by econometrics. "Success" in each and every case—and therefore the measured effectiveness of sanctions overall—is solely the result of values assigned to variables by the authors on the basis of their own personal selection and reading of often divergent case histories.\(^96\)

Compounding the problem of subjectivity (which might, in principle, be assumed randomly distributed) is the hazard of so-called "experimenter effects": the tendency of experimenters to unconsciously, yet systematically, skew observations in favor of what they believe is the hypothesis being tested. Robert Rosenthal's landmark study documents this tendency in a wide variety of contexts spanning the physical sciences, biological sciences, and behavioral sciences—on matters ranging from the observation of movements of earthworms to experiments in telekinesis.\(^97\) Abundantly documented even for the observation of seemingly "objective" facts, experimenter (or expectancy) bias is particularly

\(^96\) Moreover, the impact of subjectivity is magnified by the fact that the authors elect to multiply the outcome variable (1 to 4) by the contribution variable (1 to 4). This means that any "error" in assigning a value to, say, the contribution variable is magnified by a factor of between 1 and 4, depending on how contribution is rated. The difference between a "success" or "failure" of sanctions in the HSE model can turn entirely on the 1-point distinction they choose to make either between a "minor" and a "modest" contribution or between a "successful" and "somewhat successful" outcome.

pronounced when observers are asked to cull through masses of data and/or make judgments. As Rosenthal remarks, "It is difficult to avoid the unconscious tendency to reject for good reason data which may weaken a hypothesis while uncritically accepting those data which strengthen it." As the author elsewhere remarks, "[t]o the observing scientist, hypothesis is both friend and enemy." The specter of experimenter effects haunts the HSE study from first to last. It arises most clearly from the plain fact that the experimenters (HSE) do all of the subjective scoring, including the definition of success itself, while making no systematic effort to control for possible biases in their selection of cases, selection of case studies, selection of excerpts from case studies, or scoring of variables.

Of course, it will never be possible to eliminate all subjectivity in the study of history, and subjectivity equally plagues non-quantitative historical analyses. But quantitative studies carry the authority that numbers—with their appearance of objectivity—intrinsically command with the public. In view of that fact, scorecard compilers should take special pains to clearly acknowledge the presence of subjective judgments where they exist, document the range of plausible estimates of the various variables, and provide an analysis of the sensitivity of the results to alternative scorings of variables. And qualitative studies—case histories—should be much more carefully peer reviewed for "experimenter bids" and further steps taken to minimize bias.

4. Proxy Problems

While HSE collide headlong with subjectivity dilemmas, the BE and Sykes studies attempt to sidestep them by employing objective proxies for all independent variables (though measurement of the dependent variable, success, remains a subjective assessment of the authors). For example, the credibility of U.S. sanctions threats (an obviously relevant variable) is measured in the BE study by a dummy variable which takes the value 1 if the U.S. pre-sanction trade pattern is between $10 million

---

98. See id. at 19.
99. Id. at 25 (quoting S.S. Kety, Biochemical Theories of Schizophrenia, Part I, 129 SCIENCE 1529 (1959)).
100. Id. at 4 (quoting E.G. Boring, Newton and The Spectral Lines, 136 SCIENCE 601 (1962)).
101. See supra discussion accompanying notes 52–71.
102. See Porter, TRUST IN NUMBERS, supra note 51.
103. See Part III infra for a discussion of additional approaches to managing subjectivity problems in case studies as well as scorecards.
104. Aggressive Unilateralism, supra note 2, at 697; Constructive Unilateral Threats, supra note 2, at 318 (listing criteria as column headings in tbl.).
and $100 million, and zero otherwise. Another variable used as a proxy for credibility is whether or not the sanctions episode occurred after the September 1985 Trade Policy Action Plan or after the 1988 Competitiveness Act (both "get-tough" signals to foreign governments).

All these independent variables have the distinct advantage of being objectively measurable. The problem, however, is that none of these proxies is a particularly good measure of credibility of threat. The notion that credibility increases markedly at the point where the magnitude of pre-sanction trade reaches $10 million and then falls off sharply at trade levels over $100 million must, as hypotheses go, be considered speculative at best. So is the hypothesis that credibility increased and remained high after the 1985 Action Plan and after the 1988 Competitiveness Act: even if those trigger events brought a short-term spike in credibility of sanctions, their continuing relevance in later years is not obvious and is never demonstrated.

Proxies that do not necessarily track the variable they stand for pose an inevitable dilemma of interpretation: is the measured regression coefficient an indicator of the contribution of that variable to outcomes, or is the coefficient simply a measure of the correspondence (or lack thereof) between the proxy and the thing it supposedly approximates? The still unmet challenge in the sanctions realm is to find reliable, objective proxies for the key variables—such as threat credibility and target state political resistance to pressure—that need to be measured (or estimated) in each case. Until that challenge is met, quantitative studies will have to continue to rely on subjective scoring of ordinal factors, raising problems of subjectivity and bias.

5. The Difficulty of Defining the Temporal Boundaries and Baselines

Suppose for the sake of argument that one could find reliable measures for all relevant variables. Unfortunately, the analytical problem would still not be solved. A further difficulty arises from the problem of defining the boundaries of sanctions "episodes"—which spawns, in turn, a corollary dilemma affecting the definition of a "successful" episode.

105. The idea is that small cases (below $10 million pre-sanctions trade in the product being complained of) are less credible predicates for retaliation because the U.S. government is simply less concerned with them, while very large pre-sanctions trade flows are also less credible situations for retaliations because in those cases the U.S. will fear counter-retaliation against U.S. exports. See Aggressive Unilateralism, supra note 2, at 698–700.

106. See id. at 698–704.

107. For a discussion of the problems caused by errors in variables, see PRACTICING ECONOMETRICS, supra note 50, at 459–62.
To understand these problems and their intricate inter-relation, consider Charnovitz’s study of 14 “episodes” in which trade sanctions were threatened under the Pelly Amendment in order to induce a foreign country to comply with an international conservation agreement.\textsuperscript{108} Eight of these cases, it turns out, involved disputes over whaling activities under the International Whaling Convention, of which four separate episodes (in Charnovitz’s tally) involved Norway.\textsuperscript{109} Should these be tallied as eight cases, five cases, or one? Likewise, four of Charnovitz’s “episodes” involved disputes between the United States and Taiwan and Korea, respectively, over their failure to enforce the General Assembly drift net ban.\textsuperscript{110} Should these be tallied as four cases, two cases, or one? In the tuna-dolphin controversy (which Charnovitz omits from his database) the U.S. imposed tuna import embargoes against one or more countries on at least 34 separate occasions.\textsuperscript{111} All of these 34 embargo actions arose from a single law affecting a single issue in relation to a single problem; and the outcome of each embargo action affected the response of both sender and targets in other cases involving the tuna-dolphin regime.\textsuperscript{112} If one is concerned with measuring the batting average of U.S. sanctions or threats in artificially isolated diplomatic face-offs with individual countries, then it may make sense to treat each separate confrontation as a distinct episode. But if one is concerned, as policymakers ought to be, with understanding the contribution of sanctions (or threats) to the formation, growth and effectiveness of cooperative regimes, then it seems clear that the experience of each separate regime should be tallied as one case, with the success of trade leverage measured strategically in terms of its contribution to the formation and effectiveness of the regime overall.\textsuperscript{113}

Consider, for example, the case of the International Whaling Commission (IWC).\textsuperscript{114} It is by now well-established through case studies that

\begin{itemize}
  \item \textsuperscript{108} Analysis of the Pelly Amendment, supra note 80.
  \item \textsuperscript{109} See Aggressive Unilateralism, supra note 2, at 763–72.
  \item \textsuperscript{110} See id. at 764–73.
  \item \textsuperscript{111} See National Marine Fisheries Service, Import Prohibition of Tuna and Tuna Products, unpublished tabulation provided to the author by NMFS (available on request).
  \item \textsuperscript{112} See Use and Abuse of Trade Leverage, supra note 40, at 30–49.
  \item \textsuperscript{113} Actually, one might sub-divide the tuna-dolphin history into two cases, because U.S. objectives dramatically shifted in 1992, from inducing fishers to diligently release dolphin encircled in the process of catching tuna to a goal of ending encirclement of dolphins altogether. The latter objective was fundamentally different, and orders of magnitude more difficult to achieve than the former, for a host of reasons explained in Use and Abuse of Trade Leverage, supra note 40, at 46–51.
\end{itemize}
U.S. trade leverage played a significant role (along with world opinion and diplomatic pressure) in obtaining and enforcing, for a number of years, a world-wide moratorium on commercial whaling.\textsuperscript{115} It is also established that the moratorium has been honored by all nations (at least publicly) as regards endangered or threatened stocks, but that a few states (notably Norway, Iceland and Japan) have refused to continue to abide by the moratorium as applied to plentiful minke stocks.\textsuperscript{116} How does one score "success" in this case? Was U.S. leverage unsuccessful because it failed to deter Japan and Norway, in certain cases, from engaging in \textit{de facto} commercial whaling? Or was it successful to the extent that it reduced Japanese and Norwegian take from what it would otherwise be, persuaded them to accept observers to monitor compliance, focused their efforts on clearly non-endangered species, and deterred other states from whaling \textit{at all}? Conversely, suppose the threat of trade sanctions persuaded Norway on one occasion to abstain from all whaling, but so angered Norwegian authorities in the process that they later decided to leave the IWC and form their own whaling regime, taking other IWC members with them. Should the episode which stopped Norwegian whaling short-term but triggered a long-term fission in the regime still be called a "success"?

Consider also the case of intellectual property protection prior to negotiation of the Uruguay Round agreements.\textsuperscript{117} Suppose the United States brings two highly publicized Section 301 cases against, say, China and Korea, respectively, in response to their failure to protect intellectual property rights to the satisfaction of the United States. Let us further suppose, hypothetically, that Korea and China are defiant, but twenty other nations of significant economic interest to the United States (including all the rest of the Asian states) scurry to establish intellectual property protection regimes to head off U.S. pressure. The Hufbauer and Schott approach (and the Sykes and Bayard/Elliott scorecards) would score this as two failures, no successes—since victories won without an overt confrontation generally do not enter the database.\textsuperscript{118} Suppose next

\begin{itemize}
\item \textsuperscript{115} \textit{See id.} at 157.
\item \textsuperscript{116} \textit{See id.} at 159–63.
\item \textsuperscript{117} For a general account of the events underlying this stylized hypothetical see C. O'Neal Taylor, \textit{The Limits of Economic Power: Section 301 and the World Trade Organization Dispute Settlement System}, 30 \textit{VAND. J. TRANSNAT'L L.} 209, 225–37 (1997) (showing that the main benefit of individual Section 301 cases brought against countries which failed to pass or enforce intellectual property rights (IPR) protection laws derived, not from the success of the bilateral negotiations themselves, but from the contribution of continued bilateral trade pressure to acceptance of IPR as a suitable topic for Uruguay Round negotiation and agreement, largely on U.S. terms).
\item \textsuperscript{118} \textit{See supra} text accompanying notes 91–92.
\end{itemize}
that Korea and China capitulate, but that twenty other nations do nothing and USTR decides it lacks the resources to chase them all down. This outcome is scored as two successes, no failures. Suppose, finally, that USTR brings five Section 301 cases, fails to win meaningful concessions in any of them, but creates enough turmoil that a world-wide comprehensive agreement on intellectual property is reached in, say, the Uruguay Round. Scoring under the scorecard approach: five failures, no successes. Yet few would dispute the contribution of trade leverage to the cause of intellectual property rights protection in such a scenario.

What these examples demonstrate is that scorecard approaches focused on tactical face-offs suffer from the same deficiencies that afflict bean-counting measures of ‘enforcement’ which look only at the number of enforcement actions taken in any given year.¹¹⁹ These approaches treat all cases as equally significant and simply count the number of tactical victories. In fact, the effectiveness of enforcement is more properly measured by the degree to which overall compliance is enhanced by the enforcement actions that are taken. Likewise for sanctions: the outcomes of highly publicized individual face-offs will affect the evolution of practice, but these face-offs do not define, and their outcomes do not approximate, the contribution of sanctions to the formation and success of international regimes.

6. Model Mis-specification

A sixth challenge for scorecards and case studies alike is in identifying the right model for analysis. This entails at least three considerations: (a) choosing the “right” factors to measure; (b) finding the right functional form of the equation (i.e. specifying the inter-relationships among the variables); and (c) finding observable variables that serve as valid measures or proxies for those factors. The last consideration, measurement, is the special problem of the quantitative empiricist. The first two considerations, however, are generic to all studies of sanctions including case studies. They require that studies be informed by a clear, cogent, and reasonably comprehensive theory of behavior. One key limitation of the current scorecards and case studies alike is their lack of grounding in such a theory—a lack that reveals itself most clearly in the form of omitted variables and incorrect equations (in the case of quantitative studies) and analytically inconsistent or partial analyses in case studies.

(a) **Omitted variables.** A key element of any empirical investigation into cause and effect relationships is to identify the "independent" or "causal" variables to be considered.\(^{120}\) In the case of quantitative studies, the consequences of so-called over-specification, i.e., including extraneous variables, are relatively benign: the resulting estimate will be unbiased but less efficient (displaying a higher variance) than if a leaner specification had been chosen.\(^{121}\) Much more serious is the consequence of omitting a variable that is in fact relevant. Omitted variables that correlate with an included variable will bias the estimated coefficients of the included variable. This means they will yield "inaccurate" estimates of the degree of correlation between right-side and left-side variables—e.g. between cost of sanctions to target and sanction effectiveness. Moreover, if significance tests are applied, the inaccurate estimate may appear more "statistically significant" than it really is.\(^{122}\) In other words, the estimate will be both wrong and speciously "precise." It is important, therefore, that econometric models be based on a credible \textit{ex ante} theory that specifies a complete and plausible set of independent variables to measure.

A logical place to look for a model to use in measuring sanctions effectiveness is in International Relations (IR) theory.\(^{123}\) In recent years, political scientists and international lawyers have formed a partnership of sorts in exploring the larger question of why nations cooperate (or fail to cooperate)—an inquiry which logically subsumes the question of the role of economic leverage in producing cooperation.\(^{124}\) Although IR theory has not yet produced an explicit model for use in explaining sanctions outcomes, it has identified a number of other factors—cognitive, moral, institutional, material and individual—which work alongside, and in interaction with, leverage in shaping state behavior.\(^{125}\) A logical starting point for empirical analysis, then, would be to identify

\(^{120}\) See \textit{Practicing Econometrics}, \textit{supra} note 50, at 341–48.

\(^{121}\) See \textit{id.}, at 308–09, 312.

\(^{122}\) See \textit{id.}, at 312; William H. Greene, \textit{Econometric Analysis} 402–04 (3d ed. 1997) [hereinafter \textit{Econometric Analysis}].

\(^{123}\) For an excellent, though densely written, overview of IR theory see Andreas Hassenclaver et al., \textit{Theories of International Regimes} 134 (1997) [hereinafter \textit{International Regimes}].

\(^{124}\) This inter-disciplinary collaboration is well reviewed in Anne-Marie Slaughter et al., \textit{International Law and International Relations Theory: A New Generation of Interdisciplinary Scholarship}, 92 Am. J. Int'l L. 367 (1998).

\(^{125}\) See \textit{International Regimes}, \textit{supra} note 123; see also Oran R. Young & Gail Osherenko, \textit{Testing Theories of Regime Formation: Findings from a Large Collaborative Research Project, in Regime Theory and International Relations} 223, 241 (Volker Rittberger ed., 1993) (identifying knowledge-based, interest-based, power-based, leadership, and contextual factors).
the leverage and non-leverage ("management") factors that IR theory has found relevant to the cooperation dynamic and try to account for these factors in analyses aimed at "isolating" the role of economic leverage.

None of the existing scorecards do this, however. HSE seem to rely on native intuition in forming hypotheses and choosing variables to test in regression equations. BE announce that their selection of variables is drawn from McMillan’s game theoretic analysis of the sanctions bargaining structure. However, as will be seen below, their model does not reflect all the variables that game theory would suggest, and game theory itself (it will be seen) offers only a partial view of the forces that shape behavior in the international realm.

In any sanctions "game" there is an *ex ante* matrix that reflects for each player the "payoff" associated with cooperate and defect outcomes, respectively. The premise of sanctions is that the *ex ante* game from the standpoint of the target is a "suasion" game in which the dominant strategy for the target is to "defect" (i.e. not cooperate with the sender), even in an iterated game. The purpose of the sanctions threat is to alter the perceived payoff structure of the target so that the target state’s "cooperate" strategy dominates the "defect" strategy (at least conditionally).

From this game theory model, it follows straightforwardly that the success of a sanctions threat in accomplishing that goal, prior to imposition of sanctions, will depend on: (a) the expected economic and political impact of the sanction threatened; (b) the credibility of the threat; and (c) the target’s appraisal of the net economic and political cost of cooperating on the minimum terms sufficient to ward off sanctions. After sanctions are imposed, threat magnitude is replaced by actual economic and political impact. And threat credibility gives way to a new variable: the target’s perception of (a) the probability that sanctions will continue or expand if the target continues to resist and (b) the expected consequences of such continuance or enlargement. From a

---

126. See *Economic Sanctions Reconsidered* (1st ed.), *supra* note 1, at 33–39, where authors set forth the independent variables that comprise their "model" without reference to a single outside theory, model or authority other than their own.


128. See *International Regimes, supra* note 123, at 50–51.

129. By "net cost" I mean the cost net of any expected "intrinsic" benefits of cooperation. "Intrinsic" benefits are benefits of cooperation other than those which arise simply from the non-application or lifting of sanctions.

130. For example, I have shown that the actual economic impact of the U.S. primary embargo on tuna imports in the tuna-dolphin case was significantly enhanced by target-state fears of (a) extension of U.S. import restrictions to shrimp (in the case of Ecuador) or (b) the closure of major European tuna markets in response to continued destruction of the dolphin stocks. See *Use and Abuse of Trade Leverage, supra* note 40, at 36.
game-theoretic perspective, then, these are the variables that must be approximated in modeling the determinants of sanction effectiveness.

Neither BE nor HSE actually model these variables, however. As seen earlier, threat magnitude and credibility are poorly approximated by the BE study and completely ignored in HSE.¹³¹ Nor do these scorecards capture variations in political costs of sanctions to sender and target.¹³² In fact, these scorecards do not even provide a reliable index of economic impact, a variable which turns out to be surprisingly difficult to measure.¹³³ Most of all, current studies ignore the effect of variations

131. See supra Part II.B.1.

132. Political costs of sanctions to the target arise from the stigma associated with highly public accusations of malefeasance by another state and any tendency of those who feel the impact of sanctions to blame the government for incurring them. The significance of these costs will depend, in turn, on the degree of political influence of those who are targeted by sanctions and impute such blame—another case-specific variable. Sanctions may also, however, produce political benefits to the target to the extent that they serve as a rallying point for mobilizing public support against "foreign interference" or serve as a handy scapegoat for economic woes that arise from other sources, including the government's own mismanagement of the economy. See, e.g., Kittredge Testimony, supra note 9. Thus, the net political cost of sanctions to targets will depend on which effect predominates in a particular case an empirical question which requires close analysis of each case.

Contrary to the authors' implication, ECONOMIC SANCTIONS RECONSIDERED (2d ed.), supra note 1, at 49, HSE's distinction between "major" and "modest" policy changes does not faithfully proxy the political-cost-to-target variable, particularly inasmuch as goals as disparate (yet significant) as nuclear non-proliferation and release of dissidents appear to have been lumped in with goals such as rolling back the price of copper by two cents, under the category of "modest policy change." Id. at 41-42, 399-402.

Political costs of sanctions to sender might be expected to include the resentment of allies and tarnished reputation as a reliable trading partner, weighed against the perceived benefits of establishing the credibility of threats in future cases. It is not unusual for national security and trade policy disputes to implicate interests that have powerful cultural and symbolic resonance as well. For example, the Japan rice case (because of historic "national security" concerns and, more plausibly, the symbolism of rice in Japanese culture) raised barriers to agreement that went well beyond the economic stakes involved. See Dale E. Hathaway, AGRICULTURE AND THE GATT: REWRITING THE RULES, 20 INST. INT'L ECON. 1, 28-30, 78-81, (1987); see also James R. Moore, Unlocking the Japanese Rice Market: How Far Will the Door be Opened?, 9 TRANSNAT'L LAW. 273, 276 (1996).

133. Economic impact depends on (1) the kind of sanction chosen (for example, supplies of concessionary aid and credit are going to be harder to replace from other sources than import or export markets); (2) the scope and magnitude of the sanction; (3) the level of enforcement of sanctions; (4) the duration of the sanction; (5) the degree of target state dependence on the thing sanctioned (imports, exports, aid, credit); (6) the availability of substitute supplies of the thing sanctioned (imports, exports, aid, credit) from other foreign sources; and (7) the terms on which alternative sources of supply or markets are available. All these variables are crucially important in determining economic impact. And all, clearly, are highly case-specific. Of particular importance, and often overlooked, is the last factor—terms of replacement. As both the tuna-dolphin and South African sanctions episodes suggest, economic sanctions may have considerable impact even when alternative supplies are readily available if the cost of embargo circumvention through trans-shipment or smuggling is significant. See Use and Abuse of Trade Leverage, supra note 40, at 36 (tuna-dolphin);
in target-state-perceived cost of compliance—even though theory and common sense suggest that cost-of-compliance perceptions play a major role in target state decision-making.

I have discussed, so far, only the variables indicated by game theory, which makes a host of unrealistic simplifying assumptions: it assumes that state preferences are exogenously given, rationally determined, and fixed (apart from external inducements) in the mind of a unitary and single-minded state. But most international theorists and policy-makers recognize that these assumptions are much too restrictive: ideas, domestic and trans-national discourse, "epistemic communities," and political factions are all understood to play crucial roles in shaping each state's perceptions of what its own preferences are.\footnote{134} In practice, cognitive factors largely determine not only what state preferences are, but how strongly state preferences are held, the sacrifices that will be accepted in the interest of vindicating them, the degree of leverage that is mobilized on behalf of a sanctions threat, the credibility of the threat, the perceived costs of cooperation and/or defection by the target state, and the degree of suasion (leverage) needed to get the target state to cooperate.\footnote{135} Cognitive factors may also shape the terms of cooperation ultimately negotiated between sender and target. In fact, the very premise of the

---

EMBARGO: APARTHEID'S OIL SECRETS REVEALED 197-98 (Richard Hengeveld et al., eds., 1955) (South Africa).

Studies which omit to measure one or more of these variables potentially will miss important explanations of success or failure. For example, failure to examine level of enforcement and terms of replacement will render the analyst unable to distinguish between cases in which sanctions fail because the target state lacks sufficient unilateral leverage and those which fail because the sender state is not particularly committed to sanctions and makes little effort to enforce them (perhaps because Congress imposes them on a reluctant executive). While both situations may lead to sanction failure or diminished contribution to success, they carry very different implications for policy.

\footnote{134} "Cognitive" theory broadly encompass a range of theories focusing on the process by which state preferences are determined. For an excellent and comprehensive review of cognitive theory see INTERNATIONAL REGIMES, supra note 123, at 133–210. See also Peter M. Haas, Epistemic Communities and the Dynamics of International Environmental Cooperation, in REGIME THEORY AND INTERNATIONAL RELATIONS 168 (Volker Rittberger ed., 1993) (focusing on the role of trans-national groups of like-minded experts in shaping national preferences regarding cooperation).

\footnote{135} Of course, the influence of cognitive factors will depend not only on the beliefs of the ultimate decision-makers but on domestic political structures in the states involved. A more de-centralized governance structure will increase the number of official actors (provinces, agencies, branches of government) whose beliefs directly influence state behavior. Also, more accountable states will tend to give greater weight to public and special interest views than, say, oligarchies and dictatorships. See, e.g., Anne-Marie Slaughter, Liberal International Relations Theory and International Economic Law, 10 AM. U.J. INT’L L. & POL’Y 717, 728 (stating that one premise of "Liberalism" is that "[s]tate preferences are derivative of individual and groups preferences, but depend crucially on which individuals and groups are represented.")
USA*Engage initiative—which seeks to change Congress’s thinking about the utility of economic sanctions through argument, public pressure, and analysis—clearly recognizes the role of cognitive factors in shaping the behavior of sending states. There is no reason to assume such factors less significant at the target state level. Yet cognitive factors make no appearance in current scorecards.136

One cognitive factor of considerable importance involves what Thomas Franck calls “the power of legitimacy.”137 The idea is simple: rules or norms that are widely perceived as “legitimate” will have greater intrinsic compliance pull than rules or norms which are not so perceived.138 As a result, sanctions mobilized in support of legitimate rules and norms are more likely to succeed than sanctions whose goals are dismissed as illegitimate by the target and/or key third-party actors—other things being equal. Likewise, sanctions that are seen as “legitimate” means to an end are more likely to succeed than otherwise equivalent sanctions which are widely viewed as simple “bullying” by the sender.139 So legitimacy matters. Yet no viable measure of either goal legitimacy or means-end legitimacy appears in the HSE study.140

136. In the realm of sanctions employed to conserve the commons—e.g., efforts to protect the ozone layer, international fish stocks, endangered species—a quite different set of cognitive factors come into play. Here, the relevant factors are factors such as: the degree of consensus on the risk being addressed (ozone depletion, species extinction, etc.), the priority assigned to that risk, knowledge of alternatives to risk-causing activities, and appreciation of the costs and side-effects of such alternatives. Influencing these key cognitive variables are a range of subsidiary and contextual factors such as: the number and influence of environmental NGOs in the country in question; the size, sophistication and political influence of environmental ministries; the personal predilections of key leaders; and the nature and quality of discourse aimed at enhancing understanding of risks and risk-avoidance options. For an in-depth discussion of the nature and interaction of these diverse factors in the environmental realm, see Use and Abuse of Trade Leverage, supra note 40.


139. For example, sanctions against South Africa under apartheid and against Saddam Hussein in the immediate aftermath of the Gulf War enjoyed a much greater legitimacy worldwide than sanctions against Castro’s Cuba. As a result, more countries cooperated with Iraqi and South African sanctions and the moral and material impact on the target states was correspondingly enhanced.

140. HSE do include two scaled variables—degree of international cooperation with sender, and international assistance to target—each of which might be said to reflect, in part, the perceived legitimacy of the means and ends of sanctions. ECONOMIC SANCTIONS RECONSIDERED (2d ed.), supra note 1, at 44–45. Yet these measures fail to capture the legitimacy of the means or ends of sanctions within the target country. Moreover, cooperation indices encompass not only considerations of legitimacy but all manner of other considerations such as alliance politics, geography, and patterns of economic trade and dependence.

The closest the BE study comes to a proxy for legitimacy is their independent variable, TOUGH, which takes the value 1 if an explicit threat is issued, or if GATT ruled against the target, or if case is self-initiated; 0 otherwise. Aggressive Unilateralism, supra note 2, at 698.
As seen, omitted variables are more or less harmless if they are entirely uncorrelated with included independent variables as well as the dependent variables: omitted variables of this kind will weaken the “fit” of the regression by introducing unexplained variation, but they will not bias the estimates. A key question, then, is whether the omitted variables in the HSE study would be expected to vary with included variables, assuming reliable proxies for both could be found. In this case, prevailing theories of international cooperation strongly suggest an affirmative answer.

For example, the omitted target-state cost-of-cooperation variable quite plausibly correlates with the cost of sanctions to target and cost of sanctions to sender (two included variables). Reason: sender states will anticipate that more onerous demands require larger threats and (if threats fail) larger sanctions. Likewise, the omitted legitimacy variable is likely to correlate (though imperfectly) with international cooperation with sender, assistance to target, and cost to target (all included variables). Cognitive factors on both sides are quite plausibly related to intensity of preferences in target and sender states which will affect, in turn, the length of the sanctions episode: longer episodes being associated with both more
determined sanctioners and more determined target states. Because omitted variables are plausibly correlated with included variables in the HSE study, there is prima facie reason to believe that their "measurements" of cause and effect relationships are both biased and speciously precise.

Lesson: econometric regression is no substitute for rigorous, theoretically well-grounded explanatory models. In fact, the former requires the latter.

(b) Omitted Interaction Terms. Suppose you were investigating the causes of fire using econometric techniques. If you followed the HSE approach you would plug in separate variables representing (a) a spark, (b) dry kindling, and (c) oxygen. Assuming these variables are randomly distributed with respect to each other—that is to say, the presence of a spark does not affect the likelihood of kindling, which in turn does not affect the likelihood of oxygen being present—you might get a rather low correlation between each of these individual variables and fire. The occurrence of at least one of the above variables would result in fire only about 1 in 8 times. However, if you were to insert interaction terms into your equation—combinations of kindling-spark, spark-oxygen, and kindling-oxygen—you would get higher correlations. If you inserted a term representing the combination of spark, dry kindling, and oxygen your measured correlation would be nearly perfect. To arrive at the "right" result about the causes of fire, one has to insert a term (test a hypothesis) which reflects physical reality: fires are caused by a combination of circumstances—kindling, spark, oxygen—occurring simultaneously.

This simple example illustrates a further difficulty in the HSE and BE analyses. They failed to test for the possibility (indeed, if theory is correct, the near certainty) of interaction effects. "Successful" sanctions would seem to require the simultaneous occurrence of a combination of circumstances: in simplest terms, (a) a credible threat of (b) costs that are greater than (c) the cost of compliance with the sender's demands. The effect of a sanction costing $1 million per year depends crucially on whether the contemporaneous cost of complying with the sender demand is $500,000 or $1.5 million. The threat of a $1 million sanction linked to compliance costing $500,000 may be expected to produce quite different results, depending on whether the target believes that defying the sender will produce a 70 percent, 50 percent, or 30 percent chance of such


143. See ECONOMETRIC ANALYSIS, supra note 122, at 387–88; I am indebted to Peter Siegelman for this example.
sanctions being imposed. The key term is, in each case, the ratio of cost of sanctions to cost of compliance, multiplied by the probability of sanctions. Of course, each of these factors—credibility, magnitude of costs threatened, perceived costs of capitulation—may implicate a host of subsidiary factors which can exist in a wide variety of combinations. The point is that successful outcomes are caused by combinations and interactions of factors, including sanctions where they are used, not by any single factor acting alone. Because HSE and BE failed to take account of variable costs of compliance with sender demands, they were obviously unable to test the very plausible and important hypothesis that effectiveness depends on the ratio, perceived by the target between expected sanctions impact and costs of compliance.

7. Problems of Endogeneity

Econometric models of the kind used by HSE and BE require the implicit assumption that the "independent" variables on the right hand side of the equation \((x_1, x_2, x_3, \ldots)\) influence the "endogenous" variable \((y)\) and not the other way around. Problems of endogeneity arise when this condition is violated. The consequences are serious: biased estimates, invalid standard errors (i.e. erroneous estimates of "significance") even, possibly, the wrong sign on the independent variable—if, for example, the actual effect of \(y\) on \(x\) is greater than the other way around.

144. The credibility variable, for example, might be expected to implicate at least the following prior factors: expected cost of sanctions to sender, intensity of relative sender and target state preferences, expectation of international cooperation with sanctions. The magnitude-of-costs-threatened variable likewise implicates a host of prior variables (e.g. intensity of sender state preferences, sender state estimation of the size of a "proportional" threat in relation to the matter at issue, prior target-sender economic relations, degree of international cooperation with target/sender, perceptions of legitimacy, degree of sender state and/or international enforcement of sanctions, perceived cost of, etc.). The cost-of-compliance variable would be expected to implicate both the economic and the perceived political costs—from the target state’s perspective—if complying with the sender state demands, but appraisal of these costs would turn a range of antecedent cognitive factors of the kind described earlier.

145. To take a very obvious example, we would expect a priori that cases involving the combination of highly credible, costly-to-target sanctions aimed at achieving a relatively modest (politically or economically inexpensive) policy change would produce a much higher likelihood of success than cases involving only one of these favorable circumstances. Favorable combinations of variables produce synergies of effect that cannot be captured by simple linear models of the kind used by HSE.

146. As Baldwin has emphasized, it is necessary to adjust for the difficulty of the goal (i.e. target state costs of compliance) in each case when tallying the successes and failures of economic sanctions. ECONOMIC STATECRAFT, supra note 13, at 133.

147. See ECONOMETRIC ANALYSIS, supra note 122, at 712–14;

148. See id.
Problems of endogeneity are rife in the HSE and BE models and in the task of analyzing sanctions-backed bargaining generally. To see them, suppose we are trying to estimate the probability that, faced with a sanctions threat, the target will comply with the basic (bottom-line) demands of the sender. That probability of compliance ($P$) is going to depend, in part, on the target’s perception of the likelihood that the sender will impose sanctions if compliance is withheld, $E_P$. But $E_P$, in turn, will depend on the sender’s estimation of the likelihood that the target will give in to a threat. Call that estimate $E_P^c$ to indicate that it represents the sender’s estimate of $P_c$. Assuming there is a significant correlation between $E_P^c$ and $P_c$ and between $E_P$ and $P_s$—i.e. assuming senders and targets are not completely ignorant of each other’s expectations—we have an endogeneity problem. The probability of compliance depends in part on the credibility of the threat. But the credibility of the threat depends in part on the sender’s perception of the probability of compliance.

Nor does the problem end once sanctions are imposed, for at each decision moment after imposition of sanctions, the calculation of $E_P^c$ by targets continues, this time applied to the likelihood of sanctions persisting, expanding, or being repealed if the target remains defiant. In fact, one entirely predictable consequence of the USA*Engage reform (which calls for continuing review of sanctions and repeal of “ineffective” measures) will be to encourage future targets in the belief that their continued resistance to pressure will be rewarded by U.S. retreat. This will further diminish the credibility of sanctions threats encourage target state resistance to sanctions, and thereby fulfill the prophecy of critics who claim that sanctions are ineffective and should not be applied.149

The cat is chasing its tail, and that chase has important ramifications for both sanctions scholarship and for policy. For scholarship, it means that sophisticated techniques must be used to account for the feedback between sender state expectations and target state behavior—techniques not employed by HSE and which require certain conditions that may not be met here.150 For policy, the endogeneity problem means that public

---

149. There are further problems of endogeneity in the HSE model. For example, HSE model duration as a determinant of success rates when, in fact, success reciprocally determines the duration of sanctions: successful sanctions presumably are lifted once their goals are achieved. HSE’s “insight” that sanctions work quickly if they are going to work (Commandment Number 4), ECONOMIC SANCTIONS RECONSIDERED (1st ed.), supra note 1, at 86, may actually reflect nothing more than a statistical artifact: successful sanctions tend to be shorter-lived than unsuccessful ones, because the successful episodes cause sanctions to end sooner. Endogeneity produces, once again, a biased and speciously precise “explanation.”

150. See PRACTICING ECONOMETRICS, supra note 50, at 581–609.
analysis of the costs, impact and effectiveness of future sanctions (accompanied by a strong presumption of repeal of ineffective sanctions) is highly ill advised. It will alter the very thing being measured, and alter it, furthermore, in a negative direction. If future aggressors and autocrats know that all they have to do is survive one year of sanctions so as to generate a negative cost-benefit analysis in the next annual review, that will certainly give them a powerful incentive to hold out. Publicly announcing your cost-benefit estimate seems a lot like showing all your cards in poker—and then stating your opinion about the strength of your hand—while the wagering is still proceeding. Why would anyone in their right mind do it? This does not mean that there can be no congressional oversight of impact assessment and/or no public report. It simply means that there should be no automatic presumption that an "ineffective" assessment necessarily leads to repeal of sanctions (as opposed to alteration or escalation). And it means that the key facts and factors that underlie the ultimate decision on sanctions should be kept confidential—so that targets cannot game them. Sanctions analysis is fundamentally an intelligence estimate and ought to be treated as such.151

8. Multicollinearity

Multicollinearity measures the degree to which one or more independent variables (on the right hand side) are correlated with each other. The effect of high multicollinearity on regression estimates is to distribute variation randomly across the collinear variables, causing correlation between the dependent variable (here, sanctions effectiveness) and each of the collinear right-side variables to appear weaker than it really is.152 In the case of the HSE scorecard, the index of prior relations between sender and target, cost of sanctions to target, and presanctions trade linkage between target and sender countries and credibility of threat (where such a measure is employed) might all be expected to display a certain degree of collinearity. Since all these variables tend to move together it is hard to sort out their independent effects. Multicollinearity may provide one explanation for the fact that only two of the 18 variables employed in the HSE study are significant at the 95 percent confidence

151. Proponents of sanctions impact assessment defend the practice by analogy to environmental impact assessment, which, of course, is done publicly. See Richard Haass, Vice President and Director of Foreign Policy Studies, The Brookings Institution, Lunch Address at Symposium, Sanctions Reform? Evaluating the Economic Weapon in Asia and the World hosted by GEO. J. LAW & POL'y INT'L BUS. (Feb. 23, 2000). The analogy misses a crucial difference: the environment cannot and will not alter its behavior on the basis of what the environmental impact assessment says.

152. See PRACTICING ECONOMETRICS, supra note 50, at 435.
level. It is not hard to see that the problem of collinearity is a conceptual problem that confronts any analysis, even if it is not quantitative. It is not escaped by confining one’s efforts to case studies.

9. The Difficulty of Modeling Dynamic Processes by Static Proxies

In the simplest terms, we would expect economic sanctions to succeed if they impose (or threaten) political and economic costs that are greater than the costs of complying with the sender states’ demands. But neither the sender’s demands, nor the target’s perception of the threat, the cost of sanctions or the “cost of compliance” are static. The sender state’s demands are subject to change through at least three possible mechanisms—negotiation with the target, dialogue with other states, or internal domestic discourse. Shifting demands will, in turn, shift the target state’s “cost” of compliance with the sender’s demands. Moreover, the threat of sanctions typically spawns a domestic and trans-national dialogue within the target state that shifts the “payoff matrix” of both sender and target over time.

Likewise, the economic impact of sanctions is subject to evolution over time in response to at least four factors: (1) primary sender states may broaden or narrow the scope of sanctions over time; (2) international cooperation with sanctions may wax or wane over time in response to economic forces as well as dialogue among the lead sanctioning state(s) and actual/potential cooperators; (3) black markets may develop in sanctioned goods and services, the importance of which depends in part on the ingenuity and unscrupulousness of producers and in part on the vigor of national enforcement efforts; and (4) capital, labor

153. See Economic Sanctions Reconsidered (1st ed.), supra note 49, at 99–100. The variables, indicated by a T-statistic with an absolute value greater than 2 are: (1) use during world war and (2) the amount of support received from third countries. See id.

154. This dynamic is documented in some detail in Use and Abuse of Trade Leverage, supra note 40, at 63–73.

155. See id. at 60, 73. In the tuna-dolphin case, for example, U.S. tuna embargoes generated a dialogue about dolphin conservation options in Latin America and among Latin American and U.S. governments, fishing fleets and environmental NGOs that had the effect of fundamentally transforming both the magnitude and credibility of U.S. trade sanctions, the demands supported by such sanctions, and the target state calculation of the cost of complying with U.S. sanction demands. These sanctions-inspired cognitive shifts are documented in detail in Use and Abuse of Trade Leverage, supra note 40, at 60, 73. Similarly, the “Helms-Burton” sanctions, despite the outrage they provoked in foreign capitals, did inaugurate the dialogue about proper and improper uses of expropriated property which has led directly to a pathbreaking understanding with the European Union on investments in wrongfully expropriated property. See Sanctions Policy: Hearing Before the Senate Foreign Rel. Comm., 108th Cong. (1999) (statement of Stuart E. Eizenstat, Under Secretary of State for Economic, Business and Agricultural Affairs).
and goods markets tend to adjust to sanctions over time, usually limiting their economic impact.\footnote{156}

Finally, the probability of sanctions in time 2 (and with it the target’s anticipated cost of non-cooperation) is obviously influenced by whether or not sanctions were imposed under “similar” circumstances in time 1, and whether or not they “worked.” Indeed, game theorists have posited this inter-temporal and inter-case reputation concern as an important driver of behavior on the part of both senders and targets: targets will evaluate the probability of sanctions according to the sender’s reputation for toughness based on the past behavior of the sender; while senders evaluate the probability of success according to the target’s reputation for stubbornness based on past behavior of target.\footnote{157} The static scoring of variables in current scoreboards fails to capture this crucial dynamic aspect.

10. Invalid Cross-tabs Comparisons

HSE do not rely on their quantitative regression results exclusively, or even principally, in defending their “Nine Commandments” about the use of trade sanctions. They also rely on a practice called cross-tabulation, or “cross-tabs.” This practice, as HSE apply it, involves computing the average value of some independent variable for success cases, and the average value of the same variable for failure cases, and drawing general inferences from the comparison.\footnote{158}

For example, HSE note that the average “prior relations index” for successful cases involving a modest policy change is 2.4 (on a scale from 1 to 3), while the average prior relations index in failure cases is 2.0. From this the authors conclude: “Attack Your Allies, Not Your Adversaries.”\footnote{159} But is the difference between 2.0 and 2.3 a statistically significant difference? HSE would have us assume it is, but offer no evidence for the significance of the difference.

\footnote{156} The dynamic character of the economic impact of sanctions in the tuna-dolphin case is documented in \textit{Use and Abuse of Trade Leverage}, \textit{supra} note 40, at 34–36. HSE recognize the dynamic impact of sanctions via their assertion that the impact of foreign policy sanctions is likely to wane over time. \textit{Economic Sanctions Reconsidered} (1st ed.), \textit{supra} note 49, at 86.

\footnote{157} For a theoretical treatment of this dynamic, see Eaton & Engers, \textit{supra} note 91, at 1, 3–5.


\footnote{159} \textit{Economic Sanctions Reconsidered} (2d ed.), \textit{supra} note 1, at 99–100.
More fundamentally, HSE's use of cross-tabs—which entails drawing categorical policy conclusions about the significance of a particular variable in determining effectiveness from the average value of that variable in "successful" and "failed" cases—seems to imply that other factors are being held equal such that the effect of the variable in question has been isolated. In fact, nothing else is being held equal in a cross-tabs comparison because no regression has been performed. "Holding other factors constant" to isolate the contribution of particular variables to variations in success rates is precisely the job of multivariate regression analysis. Cross-tabs without regression leave wide open the possibility that the discrepancy in success and failure rates might be attributed to other factors entirely—such as the degree of prior trade between sender and target—or combinations of factors which may activate or nullify the effects of the variable in question. In short, drawing final conclusions from cross-tabs comparison is a very risky business. Yet virtually every one of their Nine Commandments are "supported" by just such a static, bivariate comparison.

11. Data Problems

In principle, many of the problems discussed above can be solved by examining more variables. However, the number of variables cannot be greater than the number of observations or estimation becomes technically impossible. Also, confidence intervals tend to get larger as the number of coefficients being estimated approaches the number of observations. HSE have observations of only about 100 "cases", and install at least 18 terms in their equation (not counting the needed terms that I argue that they omit). Moreover, the HSE database spans five quite dissimilar foreign policy contexts (de-stabilization, military adventure disruption, impairment of military potential, major policy change, and modest policy change) each of which arguably follows a more or less distinct political dynamic and should be modeled separately. Subdividing the data base into these five major conceptual categories would reduce the sample size even further—to an average of less than 20

160. PRACTICING ECONOMETRICS, supra note 50, at 202-04.
161. It may, of course, be true that allies are easier to sway through sanctions than adversaries. But the cross-tab comparison provided by HSE does not prove it.
162. In practice, cross-tabs are quite useful, and are widely used by econometricians, as a rough first guide to the data and as a device for generating hypotheses to test through regression. Because of the limitation reviewed above, they are not widely accepted, however, as a basis for drawing final conclusions from data. See STATISTICS, supra note 50, at 47-48.
163. ECONOMIC SANCTIONS RECONSIDERED (2d ed.), supra note 1, at 91-107.
164. See STATISTICS, supra note 50, at 493.
165. See id.
cases per category. In short, there are simply too few cases in each category—and too many variables—to permit the kind of quantitative analysis that HSE engage in.

Beyond this basic, structural problem, scorecards suffer from an additional problem arising from the quality of data in their database. The sources of data for economic sanctions scorecards are obviously either raw data or case studies of episodes in which economic sanctions have been used, or not used, to accomplish various goals. Yet, as has been seen, HSE provide no explication of the criteria they applied in choosing authors, case studies, or excerpts from case studies; do not discuss the factual foundation for the scores assigned; do not identify disagreements among the case historians upon which they draw; and do not indicate the sensitivity of their analysis to plausible alternative scorings. What is clear is that authors study cases for a wide range of purposes, only one of which is analysis of the effectiveness of sanctions. Under these circumstances, it would be a serendipity, indeed, if the HSE database of case studies turned out to be sufficiently standardized and rigorous to supply truly reliable estimates for all variables that HSE measure, not to mention the variables they omit to measure.

166. See infra Part II.B.1; Economic Sanctions Reconsidered (1st ed.), supra note 49, at 27–69 (discussion of methodology) and any of the case abstracts collected id. at 107–753 (collection of case abstracts). In fairness to the authors, providing an analysis and discussion of this kind would have added considerable length and complexity to an already formidable undertaking. So be it. Methods, assumptions and sensitivity of results to variations in assumptions need to be disclosed in any quantitative analysis that offers itself as a basis for policy, be it a risk assessment or sanctions effectiveness scorecard. For recommendations of the National Academy of Sciences see Committee on Risk Assessment of Hazardous Pollutants, National Research Council, Science and Judgment in Risk Assessment (1994).

167. At least case histories exist for foreign policy sanctions episodes. In fact, HSE have done a great service by unearthing and collecting hundreds of these case histories, which they report at the end of each case history in Economic Sanctions Reconsidered (1st ed.), supra note 1, at 107–753; Economic Sanctions Reconsidered (2nd ed.), supra note 1, at 123–298; Economic Sanctions Reconsidered: Supplemental Case Histories, supra note 65. See also case studies reported supra note 48. Trade policy and environmental policy sanctions are much less studied. The most detailed case studies to date are Duncan Brack, International Trade and the Montreal Protocol xvii (1996) (discussing the stratospheric ozone treaty), and Use and Abuse of Trade Leverage, supra note 40 (discussing tuna dolphin sanctions). David Caron and M.J. Peterson have written on international efforts to obtain and enforce a global moratorium on commercial whaling, without, however, focusing specifically and systematically on the role of economic leverage. David D. Caron, The International Whaling Commission and the North Atlantic Marine Mammal Commission: The Institutional Risks of Coercion in Consensual Structures, 89 Am. J. Int'l L. 154, 159–63 (1995); M.J. Peterson, Whalers, Cetologists, Environmentalists and the International Management of Whaling, 46 Int'l Org. 147 (1992) (discussing the moratorium on whaling). The dearth of detailed and focused empirical work on trade and environment episodes constitutes, in this author's view, a serious impediment to both scholarship and policy in the field of trade and
12. Invalid Extrapolation from Foreign Policy to Environmental Sanctions

The last problem with scorecards to be discussed in this essay arises not from the design of the scorecards themselves but from the use that has been made of them. While HSE and other compilers of scorecards have been quite scrupulous in limiting their insights to the policy realm from which their data is drawn, others have not always been so careful. In particular, the USA*Engage coalition distinguishes "high foreign policy" from commercial sanctions (discouraging the former and not the latter), but makes no distinction between foreign policy and environmental sanctions. Its publications—and its draft legislation—lump environmental sanctions into the same analytical pot as costly, ineffective foreign policy sanctions: facing an adverse presumption and burdened with special procedural requirements. So do the public pronouncements of the President’s Export Council and the National Association of Manufacturers. Prominent scholars have likewise treated environmental sanctions as if they are just a subset of foreign policy sanctions, with the same dynamic and (low) probability of success. environment. It certainly impedes any effort to draw general conclusions about the effectiveness of economic sanctions in advancing international environmental goals.

168. See Sanctions Policy Reform Act, S. 757, 106th Cong. § 1, at § 4(1)(A), (B) (1999) (including, without supporting findings, all environmentally motivated sanctions within definition of term “unilateral economic sanction” except “in a case in which the United States imposes the measure pursuant to a multilateral regime and other members of that regime have agreed to impose substantially equivalent measures”). See also Kittredge Testimony, supra note 9.

169. President’s Export Council Review, supra note 38, at 1–7 (drawing general recommendations for limiting “unilateral economic sanctions” with no analysis of environmental trade measures); Catalog of New Sanctions, supra note 4, at 6–7 (including environmentally motivated sanctions in same category as “national security” sanctions on the strength of (1) a single statement by former Secretary of State Warren Christopher that henceforth “national security” would encompass environmental protection and (2) a one-paragraph description of a single case in which the United States declined to provide export financing for hydroelectric turbines offered to China’s Three Gorges dam).

170. Chayes and Chayes, for example, infer the general ineffectiveness of treaty-based economic sanctions based on the experience of five cases: South Africa, Rhodesia, Iraq, Haiti and Yugoslavia. Three of these are cases in which sanctions were asked to achieve some of the most difficult objectives conceivable: a fundamental change of government, surrender of weapons, or (in the case of former-Yugoslavia) acquiescence to the secession of a province, Bosnia-Herzegovina. What the failure of sanctions to topple or deter hostile governments in these cases tells us about the utility of trade leverage in securing compliance with the Montreal Protocol is anybody’s guess. Yet Chayes and Chayes, far from limiting their conclusions to the realm from which their supporting data is drawn, go so far as to assert that foreign policy sanctions are likely to be more effective than environmental ones: “We derive these lessons [concerning the inefficacy of sanctions] from situations in which the UN Security Council was called upon to deal with threats to the peace, however broadly defined. But the
Yet experience and logic both suggest that the claim that environmental sanctions are "like" high foreign policy sanctions is wrong or, at best, a half-truth. The true part is that all problems in international cooperation do share certain structural and thematic similarities which IR theory explores in depth. For instance, IR theory teaches that international cooperation problems implicate a "learning" process whereby perceptions of national interest may change, either through change of mind, or through the empowerment of new groups or individuals holding certain beliefs.\(^ {171} \) International cooperative arrangements likewise tend, in varying degrees, to trigger arguments about the legitimacy and fairness of alternative rules and/or outcomes.\(^ {172} \) Cooperative problems—whether they involve nuclear non-proliferation or codfish stocks—may (or may not) involve well-developed institutions designed to reduce transactions costs of cooperation and overcome problems of collective action.\(^ {173} \) The design, legitimacy and functionality of these institutions will bear heavily on actors' calculation of the costs and benefits of cooperation. In most if not all cases considerations of power—the power of the sender in absolute terms and/or relative to that of targets—will play a role in determining both whether cooperation occurs and the terms on which it occurs.\(^ {174} \) In all cases, we might expect that the personal qualities and choices of individual leaders will contribute importantly to final results.\(^ {175} \)

All of these factors combine to determine, at any given moment in time, the political and economic "payoff matrix" (or, in laymen's terms, the costs and benefits) associated with cooperative, and non-cooperative, outcomes. At the most basic level, when economic sanctions are used or threatened, their probability of success from the sender's standpoint will depend, in all cases, on the same basic calculus: whether the sanctions threatened or imposed are large enough and likely enough to be imposed (and, if imposed, to persist) to overcome the target's resistance to coop-

\(^ {171} \) See INTERNATIONAL REGIMES, supra note 43, at 67.

\(^ {172} \) See LEGITIMACY, supra note 136; THOMAS M. FRANCK, FAIRNESS IN INTERNATIONAL LAW AND INSTITUTIONS (1995).


\(^ {174} \) For an overview of various theories concerning the role of power in international bargaining, see INTERNATIONAL REGIMES, supra note 123, at 83–135.

\(^ {175} \) See Young & Osherenko, supra note 125, at 246.
eration "on the merits." And in all cases, we would expect each of the categories of factors alluded to above—cognitive, moral, institutional, material and personal—to play significant roles in determining threat magnitude, credibility, and target state resistance.

Notwithstanding these formal similarities, however, the two kinds of sanctions are really quite different in their nature, political dynamic and mode of operation. This can be seen at the most basic level by simply reviewing some of the key explanatory variables which HSE analyze in their study of “high foreign policy” sanctions: (a) companion policies involving covert, military, or quasi-military action, (b) economic health and political stability of the target; (c) pre-sanction overall trade linkage between target and sender; (d) ratio of sender GNP to target GNP; and (e) the existence of a state of world war—in short a set of variables aimed at capturing a dynamic of one country fully mobilized in opposition to another. One would hardly expect such a study to yield much insight into the effectiveness of, say, shrimp import embargoes aimed at protecting endangered sea turtles.

Here are a few further differences between foreign policy and environmental sanctions. First, foreign policy sanctions—whether aimed at de-stabilizing foreign governments, promoting human rights, weakening an enemy’s military potential, deterring nuclear proliferation or punishing military aggression—typically implicate the fundamental security interests or the basic power structure of the target state, at least in the perception of that state. Such cases involve very high perceived political costs of compliance by the target state. By contrast, environmental sanctions seldom implicate core target state interests. Indeed, the issue in ETL episodes is often not a state action at all, but the activity of a group of private producers. In such cases, the state is not stigmatized

177. HSE, for example, examine economic sanctions having one or more of five main objectives: (1) destabilizing foreign governments; (2) disrupting military adventures; (3) impairing the military potential of an enemy state; (4) accomplishing some other “major policy change” (such as ending apartheid in South Africa); or (5) accomplishing a more “modest” policy change. ECONOMIC SANCTIONS RECONSIDERED (1st ed.), supra note 1, at 41–48. Even the so-called “modest” policy changes involve demands that are often highly sensitive in target states: e.g. protecting human rights, imprisoning or extraditing terrorists, abstaining from acquiring nuclear weapons, and returning property that has been expropriated for political reasons, as in Mexico’s nationalization of its oil wells. See id. at 41–42, 150. Any newspaper reader will appreciate that few if any of these “modest” goals actually appear modest to the target states involved.
178. The one exception may be climate change which, of course, implicates the fossil fuel energy production and consumption patterns of every state.
179. In the tuna-dolphin case, for example, the issue was the activity of tuna fishermen, Use and Abuse of Trade Leverage, supra note 40, at 13–18; in the shrimp-turtle case, the activity of shrimp trawlers, Shrimp-Turtle Panel Report, supra note 40, at ¶ 2.9–2.16.
nor is its authority challenged: indeed, the authority of the state is solicited in its capacity as a sovereign regulator of private actors under its jurisdiction. In political terms, then, the task set to environmental sanctions is typically much easier than the challenge confronting foreign policy sanctions.

Second, ETL is generally much more narrowly focused than the sweeping embargoes typically inflicted on foreign policy bugbears. A ban on imports of purse-seine-caught yellowfin tuna harvested by Mexican vessels in the Eastern Tropical Pacific is simply not the same thing, in its economic or political impact, as a wholesale ban on commercial relations with Cuba. Built into the expectations of all actors involved in the application of ETL are proportionality limits which set strict upper bounds on the degree of leverage that is likely to be exerted on behalf of the objective. In the high foreign policy realm, those upper limits are much less strict and clear. While environmental sanctions do impose economic costs on private actors in foreign lands, they inflict nothing like the misery associated with comprehensive blockades of Serbia, Iraq, Haiti, or Cuba. That powerful argument against foreign policy sanctions does not apply to the much more limited realm of environmental trade measures.

Third, ETL often (though not always) can be targeted directly at those causing the environmental harm. The ultimate targets are,

180. Such self-restraint may be variously explained by (1) rules of restraint built into the world trade law and international law generally; (2) the need to maintain order and friendly relations; (3) the need to husband the resources of leverage for use in other disputes; (4) the costliness of sanctions to the sender; (5) threats of retaliation or counter-retaliation (which will increase as the magnitude of the provocation increases); and (6) the inherent messiness, politically and conceptually, of linking unlike issues. See Howard F. Chang, An Economic Analysis of Trade Measures to Protect the Global Environment, 83 Geo. L.J. 2131, 2162-63 (1995). On international legal constraints on massive retaliation strategies, see for example, Final Act Embodying the Results of the Uruguay Round of Multilateral Trade Negotiations, opened for signature Apr. 15, 1994, WTO Doc. MTN/FA (1993), Annex 2: Understanding on Rules and Procedures Governing the Settlement of Disputes, app. 1, reprinted in 33 I.L.M. at 1244 [hereinafter WTO Dispute Understanding] (limiting trade retaliation after successful dispute settlement and failed efforts to negotiate compliance with panel or Appellate Body decision to like products as those involved in the dispute (wherever possible) and requiring equivalent suspension of trade rights in retaliation). In any case, as Chang has observed, grossly disproportionate leverage has not been an issue in the history of environmental sanctions to date, and there are good, if not conclusive, reasons to expect the pattern of self-restraint to continue. Chang, supra, at 2162-64.

181. Snake Oil, supra note 6, at C01("... economic sanctions can inflict pain on innocent people while at the same time increasing the grip of the leaders we despise. When sanctions are applied broadside as against Haiti, Cuba, and Iraq the hardest hit are the most vulnerable: the poor, the very young, the very old and the sick.").

182. For example, sanctions in the tuna-dolphin case were limited to restrictions on imports of purse-seine caught yellowfin tuna harvested in the Eastern Tropical Pacific from countries whose regulatory programs and/or fleet performance failed to meet U.S. standards.
moreover, usually economic actors. The result is an economic or at least quasi-economic calculation that is quite unlike the essentially political (and often symbolic) calculation involved in the high foreign policy setting.

Fourth, there is a heavy scientific and technical component to environmental policy making that is almost entirely lacking in most trade or foreign policy disputes. This scientific and technical dimension creates a wide scope for learning which, in turn, may either promote or impede cooperation, depending on whether the new information supports or undermines the stated goals of sanctions.\(^{183}\) Either way, the technical and scientific complexity of the ETL policy realm creates a cognitive dynamism that distinguishes the environmental realm from other areas of policy. It renders cognitive factors a uniquely important determinant of ETL necessity and/or effectiveness.

Fifth, most foreign policy and trade policy disputes involve demands for "don’ts." Don’t build atomic weapons, don’t invade Kuwait or Kosovo, don’t jail dissidents, don’t subsidize, don’t restrict imports. Not doing something places few or no demands on national administrative capacity, government budgets, or legal structures. By contrast, protecting intellectual property or the environment requires, in some cases, a very extensive government apparatus to regulate private-sector harm-causing activity. Thus, a key factor shaping the effectiveness of ETL is the capacity of the target state to implement the demands of the sender, and the degree to which that compliance capacity is enhanced by cooperative efforts mobilized by the international community.\(^{184}\) The "capacity factor," crucial in the environmental realm, is much less significant in areas where only negative demands are made.

Last but not least, environmental sanctions typically involve import sanctions, whereas foreign policy sanctions typically involve exports (and possibly imports as well).\(^ {185}\) The import focus tends to heighten the

---

The environmental offenders were fishermen and so were the targets of sanctions. See *Use and Abuse of Trade Leverage*, supra note 40, at 33–36. Compare this to the comprehensive ban on all imports, asset freeze, export ban and travel ban which the United States has imposed on Cuba since 1962–63. See *Case 60-3, U.S. v. Cuba (Castro)*, in *Economic Sanctions Reconsidered* (1st ed.), *supra* note 49, at 315–23.

183. For an illustration of the double-edged role of learning in regime formation under trade pressure, see *Use and Abuse of Trade Leverage*, supra note 40, at 21–29, 37–38, 60–73.

184. See *Chayes & Chayes*, *supra* note 43, at 197–201 (discussing the importance of capacity building).

185. Compare *GATT Legal Restraints*, *supra* note 3, at 98–106 (reviewing environmental trade sanctions actions, all of which involved import restrictions) with *Economic Sanctions Reconsidered* (2d ed.), *supra* note 1, at 59 (noting that when trade weapons are employed for high foreign policy purposes, "sender countries more frequently use export than import controls").
credibility of ETL (thereby promoting its effectiveness) because ETL is often supported by domestic interests in the sending state, who are seeking relief from "unfair" competition by foreign competitors that operate without environmental controls. But the same "Baptist-bootlegger" alliance that helps make environmental leverage more credible may also tend to make it appear protectionist and illegitimate to target state observers—thereby heightening the political costs of target-state acquiescence and tending to diminish the effectiveness of sanctions.\textsuperscript{186}

To say that environmental sanctions operate differently than trade or foreign policy sanctions in all these ways is not to say that they are necessarily more effective. As has been seen, some of the differences cut in favor of effectiveness, some cut against. But there is absolutely no basis in logic to simply assume that all these differences cancel each other out and may therefore be ignored. However tempting the grand synthesis may be, however seductive to readers and audiences, it is not valid to generalize from one broad policy context to another. As Hufbauer observed recently, "environmental sanctions occupy a 'middle space' between straight commercial disputes ... and foreign policy sanctions. Accordingly, a different template probably ought to be applied to those cases. But I haven't given serious thought to what the template should be."\textsuperscript{187} When it comes to specifying variables and interactions of variables for analysis—explaining and predicting not only whether and when but \textit{how} trade leverage works—each realm of sanctions must be examined on its own terms.

III. TOWARDS AN EMPIRICAL FOUNDATION FOR SANCTIONS POLICY

In 1993, Elliott (one of the authors of the HSE study) and Uimonen published a study plugging the relevant values for the Iraq and U.S. economies into a revised HSE-type model.\textsuperscript{188} The goal was to predict the probability that a country with Iraq's characteristics, insofar as specified by the model, would capitulate to U.S. sanctions. The results of this

\textsuperscript{186} See, e.g., Jonathan Baert Wiener, \textit{On the Political Economy of Global Environmental Regulation}, 87 \textit{GEo. L.J.} 749, 754-58, 773, 788 (1999) (noting rent-seeking tendencies of interested groups in environmental regulation generally and disparaging environmental trade measures as devices for "rent-seekers to force some inefficient regulatory regimes on unwilling victims"). For a somewhat more nuanced analysis of both the motives of sender states and target state responses to environmental trade measures, see \textit{Use and Abuse of Trade Leverage, supra} note 40, at 39–42, 73–86.

\textsuperscript{187} Gary Hufbauer, E-mail to author (Jul. 29, 1999).

\textsuperscript{188} Kimberly Ann Elliott & Peter P. Uimonen, \textit{The Effectiveness of Economic Sanctions with Application to the Case of Iraq}, 5 \textit{Japan & World Econ.} 403 (1993).
The Problem with Scorecards

analysis were by no means new, having been previously delivered to a subcommittee of the Joint Economics Committee engaged in deliberations over Iraq sanctions policy in late 1990.\footnote{Economic Sanctions Against Iraq: Hearing Before the Subcomm. on Educ. and Health of the Joint Econ. Comm., 101st Cong. 3–29 (1990) (statement of Gary C. Hufbauer, Senior Fellow, Institute for International Economics). HSE also published their optimistic prediction in the popular press. See Kimberly Elliot et al., The Big Squeeze: Why the Iraq Sanctions Will Work, WASH. POST, Dec. 9, 1990, at K1; Gary C. Hufbauer & Kimberly A. Elliot, Sanctions Will Bite—and Soon, N.Y. TIMES, Jan. 14, 1991, at A17.} No expert on Iraq (or Saddam Hussein) was invited to that hearing. The HSE model, as refined, predicted a 100 percent probability of success.\footnote{See id.}

Now, no statistical tool should be faulted for a single predictive failure. But the case illustrates three important points. First, it reveals the almost mystical authority bestowed on numbers and purveyors of numbers. Second, it reveals some of the most basic flaws of the HSE model: the disregard of such vital variables as the obstinacy of the target (huge in Hussein's case) and the international legitimacy of the sanctions (which rapidly declined, severely eroding international cooperation over time). Third, it suggests the hazards of cut-and-dried approaches to predictions of national behavior in individual cases. Scorecards, properly constructed, may be able to suggest the probability of success if sanctions of a given magnitude are applied to a country that is "like" Iraq in respect of the variables measured. But one cannot benefit from the law of averages in trying to predict the outcomes of single cases. The subcommittee would have been better off consulting an expert on Iraq and Hussein, than a scorecard on sanctions.\footnote{A recent report by the Center for Strategic and International Studies (CSIS) captured the point nicely: "Crafting targeted sanctions requires detailed intelligence about a targeted state and its elite's vulnerabilities. At present it is not clear that we have the intelligence support needed for successfully targeting sanctions over a wide set of cases." What CSIS calls intelligence, I call by the humbler nomenclature of 'data.' See Joseph J. Collins & Gabrielle D. Bowdoin, Beyond Unilateral Economic Sanctions: Better Alternatives for U.S. Foreign Policy, Executive Summary (1999) available online at <www.csis.org/pubs/beyondsancsexec.html>.}

In this article, I have not shown, or attempted to show, that economic sanctions are costless, effective, legitimate, or desirable either in general or in any particular case. Based on the wealth of case studies that have been done so far, there is, or should be, little dispute that the U.S. sanctions have often failed to accomplish their goals, and sometimes have done positive harm.\footnote{See sources collected supra note 48.} Nor have I sought to dispute the basic suggestion of USA*Engage that the costs and benefits of sanctions ought to be more carefully and systematically considered in the future. In fact, I wholeheartedly support that suggestion.
I have simply shown, first, that widely circulating reports about the large "costs" of sanctions to the U.S. economy are not supported by the study they ostensibly rely upon. Of course, costs to companies should be weighed in the balance. But there is no economic crisis here. The policy issue is not whether a given action will maximize Caterpillar's profits, but whether economic sanctions make good foreign policy, trade policy or environmental policy sense under the circumstances of each case.

Second, I have shown that we are not well equipped, at present, to deal with the effectiveness issue in any systematic and rigorous way. Anecdotes almost by definition obliterate the key contextual variables which individually and jointly determine whether sanctions succeed or fail. Case studies are more nuanced, and HSE have performed a major service in gathering references to many of these case studies together. Yet case histories by themselves tend to be complex and unwieldy. Worse, the case studies done so far are not quality-controlled, do not follow a consistent format, and do not even look at the same factors. This greatly impedes efforts to make inter-case comparisons or draw general conclusions from the experience of history. Existing scorecards of sanction effectiveness—though widely quoted and highly influential—are plagued by problems of data, measurement and method which individually and collectively undermine their validity. While some of the conclusions of their authors may make intuitive sense and may well be true, the scorecards do not prove them.

Third, I have argued that even if the current scorecards of foreign policy and trade policy sanctions were judged valid within their respective policy domains, their experience would not carry over to the international environmental realm. The failure of sanctions to topple Saddam Hussein or to persuade China to honor human rights tells us precisely nothing about the ultimate utility of trade leverage in conserving endangered sea turtles that roam the global commons. Each policy context needs to be evaluated separately, on its own terms.

Finally, on a different level, I hope I have provided the reader with a heightened sensitivity to the kinds of things that can go wrong—the things to watch out for—in quantitative analyses of social policy issues generally.

193. See supra Part I.
194. See ECONOMIC SANCTIONS RECONSIDERED (1st and 2d eds.), supra note 1, passim (citations appearing at the end of each case study); see also sources collected supra note 48.
195. See supra Part II.B.1-11.
196. See supra Part II.B.12.
What does all this mean for sanctions policy? Policy-makers are always free to conclude, of course, that the lessons of history about sanctions are clear and there is little to be gained from further study of it. "Sanctions don’t work," we are constantly told.\(^{197}\) If things are really that simple, the answer is to ban them, not study them.

Alternatively, it might be concluded that sanctions sometimes work and that "common sense" is enough to tell us when sanctions are likely to work or fail. If so, there is still no need for any new analytical endeavor. The problem with this view is that, at present, there does not seem to be any clear consensus on what that "common sense" is, either in general or in particular cases.

If, however, it is agreed that we "know" what we think we know about sanctions only by understanding past experience, and if one accepts the principal argument of this essay—that our historical analyses are not as good or reliable as they appear—then it might behoove us to pay more, and more critical, attention to the source of our "common sense" beliefs about sanctions.

While my main task in this essay has been to probe the methodological deficiencies of existing analyses of sanctions, I will close with a few constructive suggestions on ways that policy-makers might strengthen sanctions analysis in the future—if, that is, they are truly serious about analyzing the "costs" and "benefits" of sanctions prior to using or renewing them.

The recommendations that follow may be divided into two broad categories: (1) broadening and deepening the historical database; and (2) constructing a more rigorous theoretical framework for analysis. These recommendations flow directly from the critique set forth in Part II.

**Broadening and deepening the historical database.** Unearthing the true role of sanctions will require, at the outset, broadening the focus of analysis to include "quiet-threat-only" cases, "deterrent" cases, and a set of "control group" cases in which sanctions were neither threatened nor applied. Quiet-threat-only cases are relevant because they represent a significant category of successes that will be lost to diplomacy if all

---

197. *See* Kittredge Testimony, *supra* note 9, at 6 (claiming that “[t]he historical record on unilateral sanctions is clear: they do not work” while acknowledging that multilateral measures sometimes work); Pape, *supra* note 84, at 90. However, unilateral sanctions are distinguished from multilateral sanctions only by the degree of international cooperation involved. But the latter is itself an endogenous variable that depends on a host of prior factors such as the international legitimacy of the goal, the degree of third-country interdependence with the target state, the beliefs of other potential sender states about the likelihood of successful sanctions and the strength of the bandwagon effect created by unilateral sanctions that are then joined by others. So simply saying that sanctions should be "multilateral" begs a number of important, and prior, questions.
overt sanctions are abandoned. Also significant are cases where non-target states are deterred from undesirable activities, without any explicit threat, because they are unwilling to bear the costs they see inflicted upon target states who “break the rules.” These cases likewise should be tallied in any cross-cutting analysis of sanctions effectiveness. The final category of cases that needs to be more carefully considered is the baseline or “control group” of cases where no credible threat of sanctions has been mobilized, either direct or indirect. As Baldwin has repeatedly observed, the value of economic sanctions must be judged by comparing them to their alternatives.

Besides broadening the database to encompass new categories of cases, scholars and policy-makers should “deepen” the database of history from which we draw our insights. This involves three related enterprises: (1) quality-controlling the database; (2) quality-controlling the analysis; and (3) developing a consistent and rigorous template for analyzing past and future sanctions impacts. Consider first the quality of the database. Whose account of the South African sanctions story do we believe? Whose account of Iraqi or Cuban sanctions do we believe? As seen, HSE offer no discussion of either the basis for their selection of case studies, or of their selection of excerpts to quote from each study. For this I do not fault them: had they tried to do so, their study would quickly have become prohibitively long. Nonetheless, it is essential to quality control—for accuracy, reliability and completeness—the case

198. It is not clear how frequent these cases are, or how they might be found. Discovering them would seem to require eavesdropping on the councils of foreign governments. We do have agencies who make such eavesdropping their mission and it might be possible to task these agencies with preparing classified intelligence estimates of the impact of economic sanctions on the decision processes in target and similarly situated non-target states.

199. Of course, the admonition to look at threat cases is subject to caricature. U.S. statutes often authorize sanctions while leaving the executive varying degrees of discretion in deciding whether to impose sanctions in a particular case. See, e.g., Marine Mammal Protection Act of 1972, 16 U.S.C. § 1371(a)(2) (1985) (mandating a ban on imports of tuna caught with an “incidental kill or incidental serious injury of ocean mammals in excess of United States standards” but leaving the executive broad discretion in making the threshold determination of comparability). Defining “threat-only” cases to include every situation in which the United States speaks on behalf of a goal which it has the authority to support with economic sanction, would potentially sweep in virtually every U.S. diplomatic exchange! Prudence and common-sense both suggest a much narrower definition: “threat” cases are those cases in which the United States executive not only has the authority to issue a sanction but expresses a clear intention to use that authority in a well-defined manner if certain specified conditions are not met. These occasions arise much less frequently in diplomatic intercourse and, when they arise, there is no mistaking them. For an example of the distinction see Use and Abuse of Trade Leverage, supra note 40, at 20–30.

200. Economic Statecraft, supra note 13, at 123.
201. See supra Part II.B.3.
studies from which we derive the "insights" that actually shape sanctions policy.

Second, sanctions analysis must be protected to the extent possible against subjective biases. In particular, only the most Panglossian observer would assume that reliable, unbiased sanctions analysis will emerge from the current ad hoc system that pays no attention whatever to the identity and predilections of those preparing the analysis. Rosenthal's pioneering work on expectancy effects, discussed earlier in this essay, should lay this naive assumption to rest. While prior expectations cannot be entirely eliminated, it is important to minimize such effects and correct for biases that may result from them. This entails, at a minimum, ensuring that sanctions analysis is performed by scholars having a range of opinions about the utility of sanctions. It means ensuring that uncertainties and divergences of conclusions are frankly disclosed in the report of each analysis. And it means separating the analytical function from operational agencies and from those who have a financial stake in the outcome.

Third, sanctions impact analysis—to be reliable—must apply a rigorous and consistent analytical framework that reflects the best that theory and experience teach about the determinants of sanctions effectiveness. As I have shown, a number of vital determinants of the effectiveness of sanctions have been neither adequately accounted for in the HSE scorecard nor consistently scrutinized in other case studies: (1) threat magnitude and credibility; (2) economic impact of actual sanctions on target; (3) political cost of sanctions to target; (4) political cost to target of complying with sanction-backed demands; and (5) the moral compliance-pull (legitimacy and fairness) of the sender's objectives in the perception of both target and other sender states. These variables must be factored in—along with the variables with HSE do attempt to capture—in any study that purports to yield policy-relevant conclusions about sanctions.

Nor is it correct to consider these variables in isolation from each other. Sanctions of any given size can be expected to have quite different consequences depending on whether the objectives sought are large or small, intensely or lightly valued. Credible sanctions studies must

202. See supra note 97 and accompanying text.
203. See supra note 101 and accompanying text.
204. See supra Part II.B.6.
205. For example, one would expect sanctions of $41 million targeted on India and linked to agricultural reform, to have a much greater influence than sanctions of $34 million targeted at Pakistan and aimed at deterring nuclear proliferation (though both are classified as "modest policy changes."). Compare ECONOMIC SANCTIONS RECONSIDERED (1st ed.), supra note 49, at 404–08 (Case 65-2) with id. at 636–43 (Case 79-2).
capture the *relationship* between factors such as size and credibility of sanction, on one hand, and political cost of compliance, on the other. Measuring variables such as legitimacy and political cost of compliance is obviously a formidable challenge, particularly if some type of quantification is proposed. If quantification is not attempted, analysts should at least provide an ordinal sense of the rough proportionality (or lack thereof) between the size and impact of sanctions as applied and the magnitude of the objectives sought. The variables and interactions thus identified should reflect our best knowledge of all the causal pathways by which sanctions shape behavior, singly and in interaction with other factors.\(^{206}\)

Beyond specifying the correct variables, analysts must account for feedback effects between effectiveness and prior beliefs about effectiveness (endogeneity).\(^{207}\) Clearly, it is important to distinguish between sanctions which fail for objective economic or political reasons related to nature of the objective or target, and sanctions which fail to be imposed (or to win the cooperation of other states) simply because of a generalized pessimism about efficacy of sanctions.\(^{208}\) While the endpoint of these two situations is the same (failure), the policy prescriptions that flow from them are quite different.

Finally, it is vitally important to be clear about what is meant by “effectiveness” and to define the basic units of analysis (the “cases”) in a way that is consistent with that definition. As explained earlier, “effectiveness” cannot be accurately measured by counting the immediate results of individual face-offs between sender and target states.\(^{209}\) Sanctions may fail, for example, to achieve reversal of a politically motivated expropriation by a target state but nonetheless deter further expropriations by that state or others. Such sanctions achieve strategic success notwithstanding tactical failure. Likewise, sanctions may compel the target state into submission in a single face-off while arousing the redoubled opposition of other states and the rest of the world to the sender state’s demands on that issue. Such sanctions should clearly be regarded as failures. In short, sanctions outcomes therefore should be classified strategically, with “success” defined in terms of the overall progress of the international order towards acceptance of the sender’s preferences on the issue at hand, whether it be protection of intellectual property, anti-

\(^{206}\) As I have shown elsewhere, these pathways of influence can be surprisingly subtle and indirect. See *Use and Abuse of Trade Leverage*, supra note 40, at 109–10.

\(^{207}\) See discussion supra note 122 and accompanying text.

\(^{208}\) See discussion of endogeneity supra Part II.B.7.

\(^{209}\) See discussion supra Part II.B.5.