

**METHODOLOGICAL CHALLENGES IN EVALUATING THE EFFECTS OF WILDLIFE MANAGEMENT**

**REGIMES**

Ronald B. Mitchell

Department of Political Science

1284 University of Oregon

Eugene OR 97403-1284

Tel: 541-346-4880

Fax: 541-346-4860

[rmitchel@uoregon.edu](mailto:rmitchel@uoregon.edu)

<http://www.uoregon.edu/~rmitchel>

Paper presented at the

Christopher H. Browne Center for International Politics

University of Pennsylvania

04 March 2004

Draft of: 01 March 2004

NB: This is an initial effort to delineate the outlines and strategies for new research by the author in this area. Challenging criticism, whether phrased constructively or not, is particularly and sincerely welcomed.

## **Abstract**

The nature of wildlife management agreements presents analytic problems that make their influence far more difficult to assess than that of pollution control agreements. First, the variable on which information is usually available is often a poor proxy for the variable of interest, i.e., we often have good information on the number of animals killed but have relatively poor information on the number of animals still alive. Second, the temporal dynamics of wildlife management regimes makes identifying appropriate counterfactuals extremely difficult: harvest levels initially increase; it usually takes some time after the onset of a decline in harvest due to overexploitation before a regulatory agreement is negotiated; the impact of the regulatory agreement should be an even sharper initial decline in harvest to allow the stock to recover; and the “success” of the agreement lies in harvest subsequently increasing to a “sustainable” level. Third, feedbacks between reduced harvest levels, increasing prices, and changes in the effort put into harvesting present challenging factors that must be accounted for.

## **METHODOLOGICAL CHALLENGES IN EVALUATING THE EFFECTS OF WILDLIFE MANAGEMENT**

### **REGIMES**

#### **Introduction**

Treaties to regulate the use of and/or protect wildlife and other natural resources pose a set of methodological obstacles that differ from, and are arguably more difficult than, those posed by treaties that seek to reduce pollution levels. Wildlife and natural resource treaties appear to many of the same problems as pollution treaties but also pose additional problems that may prove even more challenging. These problems include determining an appropriate metric or indicator of their effects; determining if data on such indicators exists and collecting it; identifying the range of natural and non-treaty anthropogenic causes of change in those indicators; and deciphering what the effects of an agreement would look like, if there were any. This paper examines each of these issues in turn, with the goal of laying the foundation for careful analyses and comparisons of wildlife and natural resource agreements for the future.

A growing group of scholars has sought to more systematically assess and compare the effects of international environmental treaties (see, for example, Brown Weiss and Jacobson 1998; Victor, Raustiala, and Skolnikoff 1998; Mitchell and Bernauer 1998; Sprinz and Helm 1999; Miles, Underdal, Andresen, Wettestad, Skjærseth, and Carlin 2002; Mitchell 2002; Young 2003; Hovi, Sprinz, and Underdal 2003; Underdal and Young forthcoming). These efforts have begun to identify an array of different approaches for evaluating what effects environmental agreements have and identifying the sources of any observed effects.

In practice, there is increasing evidence that almost every species of wildlife that is of commercial value to humans has declined consistently and precipitously since the commencement of commercial exploitation. Most systematically, almost all commercially

viable species of fish appear to be at or below 90% of their historical population levels despite the fact that most of these stocks have been regulated by one or more international agreements or commissions (Pauly, Christensen, Dalsgaard, Froese, and Jr. 1998; Myers and Worm 2003). Whale stocks crashed dramatically from their pre-exploitation levels, despite a sequence of international agreements to stem that tide. Elephants, rhinos, and tigers are just three of the endangered species that have continued to decline in most, if not all, of their natural range despite the efforts of the Convention on International Trade in Endangered Species (CITES). Fortunately, there are some success stories, including fur seals and polar bears that went through dramatic declines in their populations prior to international regulatory agreements being established, and those agreements appear to have contributed to improvements in the plight of these species. This variation in the fortunes of different species provides both hope and opportunity -- hope that such declines are not inevitable and the opportunity to conduct research in a way that identifies the conditions and mechanisms by which intentional human action can help reduce the likelihood of such declines in the future.

This paper attempts an initial exploration of the unique difficulties of evaluating the effects of wildlife treaties, contrasting evaluation of their effects with those of pollution treaties. The paper begins by discussing the virtues and difficulties of using different dependent variables as indicators of the effects of such treaties and then discusses how variation in those variables over time could be used to identify such effects, if they were to exist. The paper goes on to discuss the difficulties of identifying independent variables other than the treaty that can be used to ensure that changes in whatever dependent variable are used are not spuriously attributed to the treaty, identifying in the process the difficulty of finding variables that make sense for use across a range of wildlife agreements. Analytic techniques are then discussed, focusing on the

unique opportunities with some wildlife agreements, most notably fisheries, to compare multiple similar international legal arrangements to identify what elements of those agreements explain variation in the effects of different treaty structures. Some complicating difficulties are discussed prior to a brief conclusion.

### **Selecting an appropriate dependent variable**

Consider international agreements that address the protection -- and regulation of the killing of -- whales, polar bears, innumerable fish species and stocks, various migratory birds, or endangered species. In trying to identify the effects of such agreements, an initial obstacle lies in the choice between using data regarding population or data regarding harvest or "take" levels. Consider the implications of each of these two alternatives.

### **Population levels as indicator**

Using population levels as an indicator for agreement effects has some advantages and some disadvantages. Certainly for those agreements which explicitly identify protecting the population of a particular species or protecting a species from extinction, population would appear to be the appropriate goal, since it corresponds with the explicit goal the member countries established for themselves and to which they agreed. In many cases, the negotiators quite explicitly link the need for stock conservation and rational management or optimal utilization of some resource, recognizing that the latter depends on the former. Using objectives delineated in agreements makes sense as the appropriate metric since, although agreements may fail to have positive effects with respect to those goals agreed to by the member countries, they are even less likely to have such effects with respect to those goals to which member countries did not agree.

Equally important, reliable population figures can be interpreted quite unambiguously. Known increases in population levels are clearly evidence that progress is being made toward the goals of the agreement. Determining whether such progress is due to the agreement or other factors requires further interpretation, but determining whether an increase in population reflects a desirable or undesirable circumstance is not. As discussed below, the same assumption cannot be made for harvest.

There are several downsides to using population information. The first is simply data availability and reliability. For many, and perhaps most, wildlife and natural resource treaties, appropriate levels of data are not available. For many species, particularly fish, whales, and other marine species, population figures are not available or are not available over the relevant time periods. In many cases, estimates of population may not commence until after the agreement is signed. When estimates of population are available, they are often unreliable (being based on a variety of contestable modeling assumptions) and, even more frequently, are not available at the level of temporal or spatial resolution that would be necessary for useful analysis. Estimates may be available at a global scale and at ten year intervals, making them of little use for evaluating the effects of an agreement adopted in a particular year that targets the stock of a species in a particular area.

A second important obstacle is that the populations of most species are driven by two sets of poorly understood factors other than international agreements: natural factors and non-treaty anthropogenic factors. Most populations of most species have ebbed and flowed over historical time long before and continuing since the arrival of humans. For most species, the magnitude and periodicity of these natural variations are not fully understood. Equally important, population dynamics are likely in many cases to depend on the pre-existing level of the species --

as a stock level declines, the magnitude and periodicity of stock fluctuations are also likely to change but in ways that make accurately estimating what populations would have been without an agreement particularly difficult.

Beyond simple natural variation, humans are influencing most animal stocks through a variety of non-treaty pathways. Consider, for example, the myriad fishery agreements that are in existence. Most such agreements regulate human extraction through the activities of relevant fisheries, and that is surely a major influence on populations. That said, stock levels are also influenced, to varying degrees, by other human activities. Habitat encroachment or destruction has dramatic influences on populations of many land animals, such as polar bears, elephants, rhinos, and tigers. But it can even have important implications for many fisheries. Consider, for example, how stocks of anadromous fish (that reproduce in freshwater but live most of their lives in saltwater) have been dramatically influenced by the extensive worldwide damming of rivers. Other human activities can also have direct effects on populations of various species, whether through pollution levels, temperature changes, or the introduction of invasive and alien species that are either competitors or predators. Thus, for many but not all species, humans may have quite direct impacts on their populations through behaviors that are unrelated to those regulated by an agreement.

Equally important, there are human activities that may wield an indirect influence on populations of various species. By indirect, I mean human activities that cause change in a related species which, in turn, have effects for the species addressed by the agreement. Essentially, this point simply recognizes the interconnections among species. Consider, for example, how human-induced changes in seal populations due to effective or ineffective controls on harvests, would influence the population of polar bears that feed on them. In these and

similar cases, there may be two, three, or four links in the causal chain between human activities in one arena and their impacts on the population of the species being regulated. The increasing number of causal links means that determining the effects of such human activities may be more difficult to identify, but it does not make it any less real. That is, we may be able to know, unambiguously, that those activities influence the species of interest but be completely uncertain as to how and how much.

### **Harvest levels as indicator**

Using harvest, kill, take, or the various other names used for direct human extraction of a species as the indicator also has advantages as well as disadvantages. Perhaps the largest benefit of using harvest levels is that harvest is more closely connected in a causal sense to agreement influence. That is, if an agreement has any influence on harvest, those effects are likely to be relatively directly traceable and evident in relatively direct changes in harvest. The effects of an agreement that actually reduces the hunting of elephants or the harvest of salmon should be visible in relatively straightforward and, at least potentially visible, alterations in the context of incentives, opportunities, information, or values expected to influence the hunters' or fishers' behaviors. Certainly a range of other factors influence harvest levels, as described below. Indeed, precisely because actual harvest levels of most wildlife depends on both effort put into harvesting and on the population available for harvesting, harvest levels are in part driven by all the factors just discussed. Even thinking through this process suggest that the more appropriate indicator of agreement influence is harvest effort, e.g., number of vessel hours spent fishing, number of hunting days in the season times number of hunters, or similar metrics. Unfortunately, such metrics are often hard to come by. Data on harvest has the advantage over that on population and harvest effort of being available in relatively accurate data at relatively high

spatial and temporal resolution. Certainly, reporting inaccuracies are possible in almost all cases. But inaccurate reporting is often averted due to other social factors and constraints having to do either with norms against lying; the collection of such statistics for domestic economic reasons coupled with the difficulties of keeping "two sets of books"; or the collection of such statistics as part of transactions in which the hunter or fisherman may have incentives to report inaccurately but the person purchasing their harvest has no such incentives.

Perhaps the most troubling problem of using harvest as an indicator of agreement effect is the fact that changes in harvest levels can indicate either the positive effect OR the lack of any effect of the agreement. Consider that, it is likely that harvest is declining prior to any agreement being signed -- indeed, declining harvest is, in most cases likely to be a necessary condition for agreement negotiation since absence such a decline the incentives to avert overexploitation are unlikely to be sufficiently strong to lead countries to enter negotiations. Once an agreement takes effect, how should we then interpret a subsequent decline in harvests? On the one hand, it can indicate that the agreement has led fishermen or hunters to constrain their behavior. But an equally compelling argument can often be made that the subsequent decline (and harvest levels below set quotas) simply reflects the fact that overexploitation has left yet fewer fish or animals for hunters or fishermen to catch. Indeed, consider Figure 1, which depicts the quotas and harvest levels for whales under the International Convention for the Regulation of Whaling of 1946. An initial reading of this chart might lead to the conclusion that compliance with the rules was quite good, since harvests in each year are below quotas for those years. However, an equally plausible interpretation of the chart can be derived by noting that, beginning with the 1962-1963 season, the quota in any given year simply reflected the catch from the previous year. Thus, the same set of facts can be interpreted either as the International Whaling Commission

(IWC) constraining, in however limited a fashion, the pressures for whaling or, alternatively, as the IWC simply presiding over the continued overexploitation of various species of whales. And, more generally, declines in harvests of any species after its regulation almost always can indicate either compliance with those regulations or a continuing decline in the underlying population from which animals are being harvested.

Another difficulty in using harvest data is that, the very effort at regulation may well alter the reliability of the data available. In a process that is a complement to the ways in which "increased policing" can lead to increases in *reported* crime rates even while reducing the overall crime rate (reported plus unreported crimes), new regulations can lead information that was freely reported previously to be driven from the system. So long as the regulatory system relies on sanctions, whether explicitly called for or more implicit and diffuse social sanctioning, regulation can reduce the availability and accuracy of information on harvest out of system. As evident in response to the Montreal Protocol regime on ozone depleting substances, previously legal behaviors which were relatively easy to track increasingly transformed into clandestine activities which were increasingly difficult to identify accurately (Clapp 1997).

### **Catch per unit effort as indicator**

An alternative indicator that provides some analytic leverage on the problem is "catch per unit effort," i.e., the number of animals killed (or weight, for most fish species) divided by the amount of time or resources spent to kill that many animals. CPUE has two primary virtues as an indicator of the effects of a wildlife regime: data on CPUE correlates reasonably well with population but is often more readily available and more accurate. CPUE correlates with population for obvious reasons -- using a given technology, a given amount of effort is likely to produce more animals caught from a larger population than a smaller one. Thus, CPUE goes up

as population goes up and goes down as population goes down. Of course, whether that relationship is directly proportional or is positively correlated in more complex ways depends on a variety of factors related to both the species being hunted and the technology being used. The fact that technologies related to hunting or catching various species changes over time must be taken into account when attempting to interpret CPUE data as an indicator of the rise or decline of a species' population.

CPUE figures are, as with harvest, also susceptible to reporting errors and intentional misreporting. Equally important, there are a range of factors that influence CPUE, many of which are not regularly or consistently reported in ways that would make CPUE figures truly comparable. Thus, two fishing boats could each spend an equal amount of time fishing and return with quite different catch levels, dependent on the number and skill of the crew, the size of the boat, the weather conditions, etc. Although some of these factors may "even out" across reports (e.g., weather is unlikely to bias estimates of population in one direction or the other), the affect of other factors are likely to change over time or across observations in ways that have a systematic influence on this indicator but in a direction that is unknown.

These and other factors mean that catch per unit effort is often not available in reliable datasets covering long periods of time. That said, when consistent, reliable, and interpretable CPUE data is available, it may prove quite useful in contributing to an effort to determine whether a regime has been influential or not.

### **Interpreting the data**

Before spending considerable time and resources collecting data, it proves useful to consider how we would interpret a graph assuming we gained access to "perfect" data. Assume,

initially, that one has both population and harvest data. What sorts of graphs would one expect from such data, if an agreement were effective? To address this question, consider the "idealized facts" regarding population trends over time. It seems plausible that those facts would be something like the following:

- Before human exploitation: Prior to human exploitation, presumably, there is some dynamic equilibrium state in which a species population rises and falls in response to "natural," non-human processes. A large fish stock might increase or decrease by as much as 5% per year from a long-term historical average, for example.
- Natural system compensates for human exploitation: During an initial phase of human exploitation, that exploitation is at a sufficiently low level that it does not effect the population of the species. In this period, changes in population due to natural factors are sufficiently large that human exploitation has no observable impact on the species population. Thus, for a fish stock with "natural" variation of 5% per year, initial levels of exploitation at one tenth or one quarter of one percent per year might be absorbed by the far larger natural variation, with the natural system adapting to human withdrawals from the system in ways that maintain populations within the same band of natural variation.
- Harvest impedes natural recovery of population's historical average: During the next phase, human levels of exploitation increase to the point that they begin to match the magnitude of natural variation, causing initial declines in population. Thus, exploitation levels of 2 or 3% of the fish stock per year might exceed the ability of the natural system to respond, with population experiencing a systematic bias toward lower populations and with recovery above or to the historical average remaining possible in some years but becoming increasingly infrequent.

- Increasing harvest/decreasing population: During a third phase, population continues declines even as harvest increases. Thus, a 2% (of initial stock) exploitation rate can easily become 4% and then 6% and then 8%, as technology or number of actors exploiting the resource increases. These increases lead to higher harvest levels, which are increasing (though at a decreasing rate of 100%, then 50%, then 33% in the example, though other rate trajectories are quite possible) but declining population as exploitation exceeds not only the natural systems ability to compensate for human exploitation but such exploitation prevents the natural system from replacing itself, rather than simply predisposing it against doing so. This increasing harvest/decreasing population phase can continue for an extended period of time. Indeed, this divergence between harvest and population provides the foundation for understanding why those dependent on a wildlife resource often can either fail to see or ignore evidence of declining population.
- Decreasing population/decreasing harvest: During a fourth phase, harvests and CPUE begin to decline dramatically enough that it becomes increasingly difficult for harvesters to ignore these declines or attribute them to natural variation in the population. This recognition may take an extended period of time but, eventually, is likely to lead for calls -- by at least some harvesters -- for collective action to remedy the situation. In short, the harvesters recognize they have instantiated a "tragedy of the commons" (Hardin 1968) in which they must establish "mutual restraint, mutually agreed upon;" must turn to third parties for regulatory assistance; must find some other solution to the problem (e.g., privatization); or they must accept the likely outcome of collective destruction of the resource on which they depend.

- Regime formation: At the international level, actors that succeed in negotiating a wildlife regime usually seek, in some fashion or other, to reduce harvest in the hopes of reducing the pressure on the species, either cutting harvest sharply so that populations can recover more quickly to a level at which more sustainable harvests can be caught, or limiting harvests less sharply in hopes that such constraint will allow the harvesting industry to remain viable in the short to medium term while improving their long term future prospects.
- Reduced harvest period: In some initial period after the regime is formed, we expect harvests to decline whether or not the regime is effective. If the regime is being effective, presumably, harvests drop quite dramatically. Recall that harvests have already been declining prior to establishment of the regime: indeed, recall that declining harvests are the impetus for regime formation. Therefore, one major analytic obstacle arises in attempting to identify a plausible counterfactual for the decline in harvests that would have occurred absent the agreement relative to the decline in harvests observed with the regime. With, at least some, pollution agreements, it is quite plausible to argue that, absent a regime, pollution levels would continue to increase at historical rates of increase. By contrast, for a wildlife regime to be effective, it must respond to a decline in harvest levels that is projected to be long-term by inducing an even sharper, but shorter-term decline in harvest levels. Thus, counterfactual creation must be far more precise because it must generate counterfactuals that have a resolution beyond simply getting the sign right but also getting the magnitude sufficiently right to distinguish the regime's effect, if any.

- Recovered harvest levels period: Wildlife regimes divide into two types, those that seek to recover stocks sufficiently to allow a relatively high-magnitude and ongoing human exploitation of the stock (usually referred to as conservation) and those that seek to essentially ban harvest altogether (usually referred to as preservation). For conservation regimes to be successful, they must reduce harvests in the short to medium term, allowing populations to recover, and then allowing harvest levels to rise again to "sustainable" levels (whether maximum sustainable yield or not) that balance human exploitation against natural replenishment rates. For preservation regimes to be successful, they must reduce harvests more permanently.

Figure 1 reflects a stylized version of these phases in the life cycle of harvest of a species in response to a successful wildlife agreement and demonstrate that the relatively straightforward techniques for assessing regime effects seem far less available than for pollution treaties, as illustrated in Figure 2. In pollution regimes, a reasonable argument can be made that the impact of the treaty is evident simply in the change in the slope of the use of pollutants. That is, a pollution agreement is successful in a counterfactual sense if it causes, *ceteris paribus*, the changes in rates of emissions after its enactment to be lower than those changes in rates of emissions were before its enactment. If rates of emissions were increasing prior to the agreement, counterfactual success consists simply of ensuring that, subsequently, those rates, even if not declining, are increasing at a slower rate.

By contrast, the trajectories of counterfactually successful wildlife agreements, are more complex and more difficult to interpret. Many, and perhaps most, wildlife agreements do not establish a temporary moratorium or ban on harvest but instead impose slowly declining quotas, due to their simultaneous commitment to preserving the relevant harvesting industry (which

often could not survive a moratorium or strict regulations that may need to last five or more years) and due to economic and political pressures to ignore scientific advice that may suggest quite sharp harvest limits that are economically very costly. It is also difficult to know how long the period of reduced harvest or moratorium must last to be effective ecologically. That is, a treaty with which relevant parties complied completely might still fail to re-achieve high sustainable yields if harvests were not sufficiently restricted during the period of sharp treaty regulation.

### **Independent variables**

Environmental treaty assessment has usually been approached by examining available evidence to see whether that evidence is consistent with the effect of the relevant treaty. Here, I adopt the converse approach of identifying whether a treaty has been influential by attempting to find the sources of variation in the harvest trajectory for some regulated species and then determining whether the treaty is among them. In some sense, the analytic enterprise then looks less like painting and more like sculpture, attempting to remove (or control for) the influence of non-treaty variables on the levels of harvest and, after having done that for the "most likely suspects," then assessing whether there remains some unexplained variation in that harvest level that is consistent with the impact of the treaty.

Thus, the first element of assessing a treaty's influence on harvest involves identifying the other factors that also are likely to influence harvest of that particular species. On the one hand, one can imagine a generic set of factors being relatively well-correlated with a broad range of pollutants. That is, variation (whether over time or across countries) in such factors as population and economic activity are likely to do reasonably well at explaining a significant

amount of the variation in sulfur and nitrogen oxide emissions, CFC emissions, discharges of salts and heavy metals into rivers and lakes, and levels of hazardous waste. Equally important, aggregation to the national level does not seem likely to undercut whatever correlation exists significantly -- that is, since one averages both high economic growth and high emission levels in the urban area of a country, for example, with low levels of both for rural areas, correlations across countries seem likely still to be evident (even if the strength of their correlation is diminished).

On the other hand, it seems at least initially likely that the factors that are best correlated with population and harvest of particular species are likely to be quite particular to that species, not only in terms of the factors involved but also in terms of their regional specificity. Changes in the population of a species are likely to be affected by human forces beyond direct harvest, including habitat loss and non-harvest loss due to humans. Thus, polar bear populations and harvest are influenced both by declining habitat due to direct encroachment by settlements as well as by declining levels of sea ice that increasingly appear to be early evidence of climate change. At the same time that settlements reduce the habitat available to polar bear populations, they also bring polar bears into more frequent contact with humans and human waste products. Thus, polar bears are more likely to be killed in self-defense as humans and polar bears co-habit in the same regions, and they are also more likely to be poisoned by accidentally ingesting chemicals and other products of human habitation. Notice that such factors are quite place-specific, the average population of a country may stay relatively constant during the same period that the encroachment of humans on particular habitats may change dramatically.

Other factors that are likely to influence wildlife harvest include, as already mentioned, the price for the species in question, the number of harvesters engaged in the industry, and the

technology or technologies being used in the process of harvest. The precise levels of these factors are probably less important than relative changes in these levels over time or variation in these levels across different countries or other units of analysis used in any study.

### **Analytic techniques**

Whether adopting qualitative or quantitative methods to assess the impact of a wildlife regime, the analyst must somehow compare the set of behaviors observed with the regime in place and compare them to an estimate of the set of behaviors expected but unobserved had the regime not been in place (King, Keohane, and Verba 1994; Fearon 1991; Biersteker 1993). The foregoing section has argued that the predicted trajectory for catch if a regime is successful is quite complex, declining sharply directly after the regime, slowly recovering, and then leveling off subsequently. One, perhaps surprising, analytic virtue of this complex trajectory is that its very complexity makes it less likely that, if such a pattern of behavior actually occurs it would be due to some other factor or set of factors. That is, regime success produces a trajectory unlikely to be replicated by other factors and, so, if we observe such a trajectory it is unlikely to be due to other factors. However, complete regime success is likely in few, if any, regimes. The forces that create a suboptimal situation to which states feel a need to respond, the anarchic realm within which states must operate, and a variety of other factors provide strong arguments for the case that regime success is a rare event. In the present context, this implies that the observed trajectory of even a "successful" regime is unlikely to appear similar to a fully successful regime and, instead, is likely to be ambiguous in its interpretation. Thus, carefully estimating counterfactuals against which observed outcomes can be compared will be crucial to any convincing argument regarding regime effects.

What empirical evidence can be used to generate plausible counterfactuals? In evaluating wildlife regimes, several opportunities may present themselves.

The first source of evidence of "what would have happened otherwise" comes from comparing the behavior of member countries after a regime is formed (or their behavior after they join a preexisting regime) to that of nonmembers during that same period. Such a comparison has the benefit of holding constant an array of time-varying, individual-invariant variables, i.e., factors that vary over time but the variation in which affects all countries equally, such as changes in world prices for the species in question (assuming there is a world market for the species). Evidence that the harvest of nonmembers has not declined as sharply as that of members can demonstrate that the decline in harvest of members is due to restraint rather than a decline in the availability of the species. Thus, reports that the Soviets had been killing blue and humpback whales but not reporting such kills to the International Whaling Commission, however troubling, does, after the fact, provide evidence that supports the contention that the failure of Iceland, Japan, Norway, and other whaling countries to kill such whales during the same period was due to real restraint rather than because the numbers of such whales made them impossible to hunt successfully. In short, the behavior of nonmembers may allow the analyst to distinguish whether low member harvests are the result of member restraint or population scarcity.

The second source of counterfactual evidence comes from comparing the behavior of countries before they join the regime to their behavior after they join. The decline in harvest over time (i.e., the first derivative), and the decline in the decline in harvest (i.e., the second derivative), provides some evidence of the "expectable" future trajectory of harvest for that country. To the extent that catch declined steadily for some period prior to treaty formation or to

the country joining the treaty, one might expect catch to continue declining at that same rate. A sharper decline provides evidence consistent with an argument that the decline is due to self-conscious restraint rather than the continuation of a preexisting decline in population available for harvest. That said, a sharp decline in harvest may still arise due to a non-linear response of the population to overexploitation.

A third source of counterfactual evidence comes from comparing member behaviors with respect to the species or stocks regulated by a treaty to their same behavior with respect to similar species and stocks not so regulated. The regional approach that has been taken to international fisheries management provides a rich set of cases for comparison in this regard. Whereas most environmental problems have been addressed either through global regimes or through regional regimes that have not been replicated in other regions, a broad array of fisheries regimes have been established to address different species or stocks, in different ocean regions or rivers, and involving different countries. Thus, the Convention for the Conservation of Salmon in the North Atlantic Ocean (NASCO), the Convention for the Conservation of Southern Bluefin Tuna, and the Agreement for the Establishment of the Indian Ocean Tuna Commission all regulate harvest of particular species of fish in delimited ocean regions. Yet, in each of these cases, the species regulated are harvested within the regulated regions by countries that are not members of these treaties, migrate outside the regulated region where even members can harvest them legally, and are species that have close substitutes in the international marketplace.

Essentially, careful development of datasets can allow comparisons such as:

- Harvest by a member country of a regulated species within a regulated region to harvest by that same country of that same species outside the regulated region; or

- Harvest by a member country of a regulated species within a regulated region to harvest by that same country of a similar, non-regulated, species within the regulated region.

Since fisheries management regimes frequently have regulated by placing constraints on the fishing gear that can be used for harvest, comparisons can also be made between harvest by a member country of a regulated species within a regulated region with a given type of gear to harvest by that same country of that same species with the regulated region but with a different type of gear.

Arenas such as fisheries -- and others in which regimes regulate only some fraction of a relevant wildlife harvest or regulate one species but not a similar species that is an effective economic substitute -- provide opportunities for comparing regulated behaviors to non-regulated behaviors and thereby making relatively compelling and convincing judgments about the likely behavior counterfactual to that observed by regulated actors.

Of course, many wildlife regimes are global in scope and regulate all actors involved in the harvest of a given species. Thus, the Convention on International Trade in Endangered Species (CITES) regulates trade in elephants, rhinos, tigers, and a wide variety of other threatened species, the International Convention for the Regulation of Whaling regulates worldwide whaling, and the Agreement on Conservation of Polar Bears, although involving only five countries, includes all countries needed to account for the full range of polar bear habitat. In these cases, the opportunities for developing counterfactuals are more limited.

Arenas in which multiple regimes regulate relatively similar species may provide unique opportunities for making claims about relative effects of regimes (Mitchell forthcoming). In most cases in international relations, there are only one or two regimes regulating an issue area and so it proves relatively difficult to draw convincing generalizations across cases (because

such generalizations are based on too few cases to be convincing) and make it difficult to identify what specific features of successful regimes are the source of their success. This latter proves a particularly difficult problem but one which is centrally important to policymakers seeking to design more effective international agreements. Consider an international regulatory agreement that some scholar has shown convincingly to be successful: that agreement undoubtedly has a wide variety of features, any one of which might be convincingly argued as responsible for its success. Equally important, that agreement was undoubtedly implemented within a context that may have consisted of a set of conditions that greatly facilitated that success, in some ways making that success less impressive because conditions made it a "most likely case" for success. Careful process tracing and use of counterfactuals may be able to help eliminate many of these alternative explanations, leaving a relatively few standing (George and Bennett forthcoming). However, even so, empirical evidence from a broader array of cases could be particularly valuable in providing stronger evidence that certain agreement features have consistent effects across a wide variety of contexts and other agreement features are relatively unimportant to an agreement's success. Since theories already exist that support the inclusion of a wide range of agreement features (from sanctions to rewards to transparency and accountability), and since, post hoc, it is quite easy for scholars to identify a theory to explain why some features previously expected to be unimportant nonetheless proves central to a given agreement's success, having a set of comparative cases that vary both with respect to their features and the contexts within which they are implemented can be a valuable tool in responding to this embarrassment of theoretical riches faced in the realm of international regime effectiveness research.

### **Complicating issues**

Beyond those issues already raised, several other factors make evaluating wildlife treaties more complicated than evaluating pollution control treaties.

One important one is the difficulty of addressing the relationship between the market price of the regulated species and the quantity produced. As already delineated above, we have good reason to believe that an important driver of the level of harvest of a species is the market price for that species. But, because of the relationship between regulation, market supply of the species, and market price, price cannot be taken as strictly exogenous to the regulatory process. That is, price is, at least in part, dependent rather than independent of international regulation. To the extent that a regulatory regime succeeds at artificially reducing harvest, it influences the industry in a fashion similar to that of the industry itself forming a cartel. By artificially reducing market supply, the regime increases the market price for such supply as does get to market. That increased price, however, provides incentives for both those economic agents already in the market to forego whatever restraint they had previously shown as well as for new agents to enter this market. Pollution is produced as a byproduct of economic activities undertaken for other reasons -- reducing pollution usually entails imposing additional costs on the economic activity that produces it. By contrast, reducing harvest of a marketable species has the perverse effect of actually increasing the market price and thereby creating pressure for reversion to the pre-regime equilibrium level of harvest. This negative feedback loop has two implications for analyzing regime effects. The first is to make it ambiguous as to whether (and how) price should be used in any model of harvest. As with the ambiguity regarding the imputation of population figures from harvest data, so too with price and regulatory influence. In most cases, assuming other factors can be reasonably assumed to have remained relatively

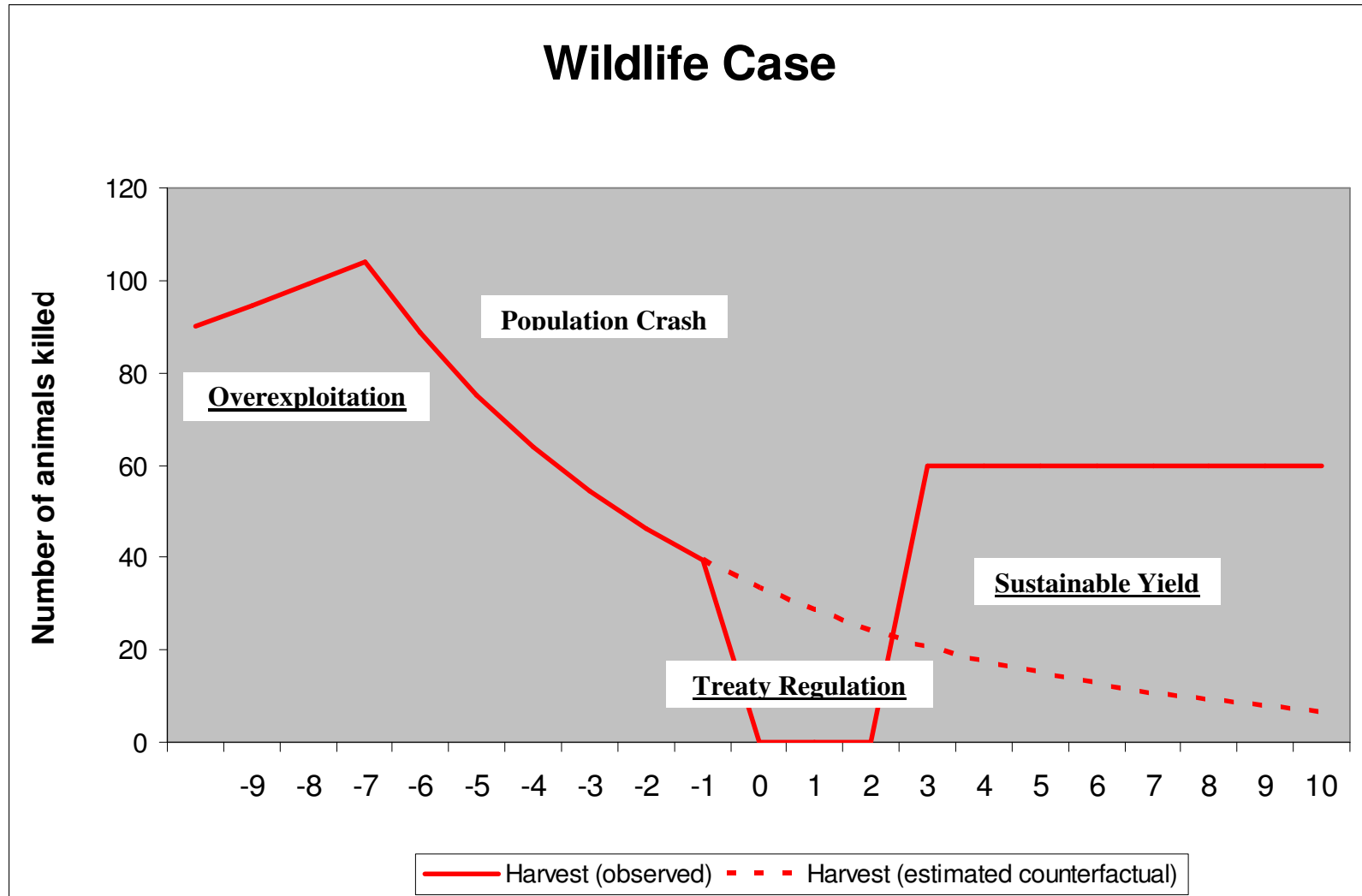
constant, price increases will reflect reductions in market supply. However, it will be difficult in many cases to determine whether that reduction in market supply results from effective regulation that has led to harvester restraint or from ineffective regulation that has allowed continued overexploitation, a corresponding decline in population, and therefore a decline in harvest.

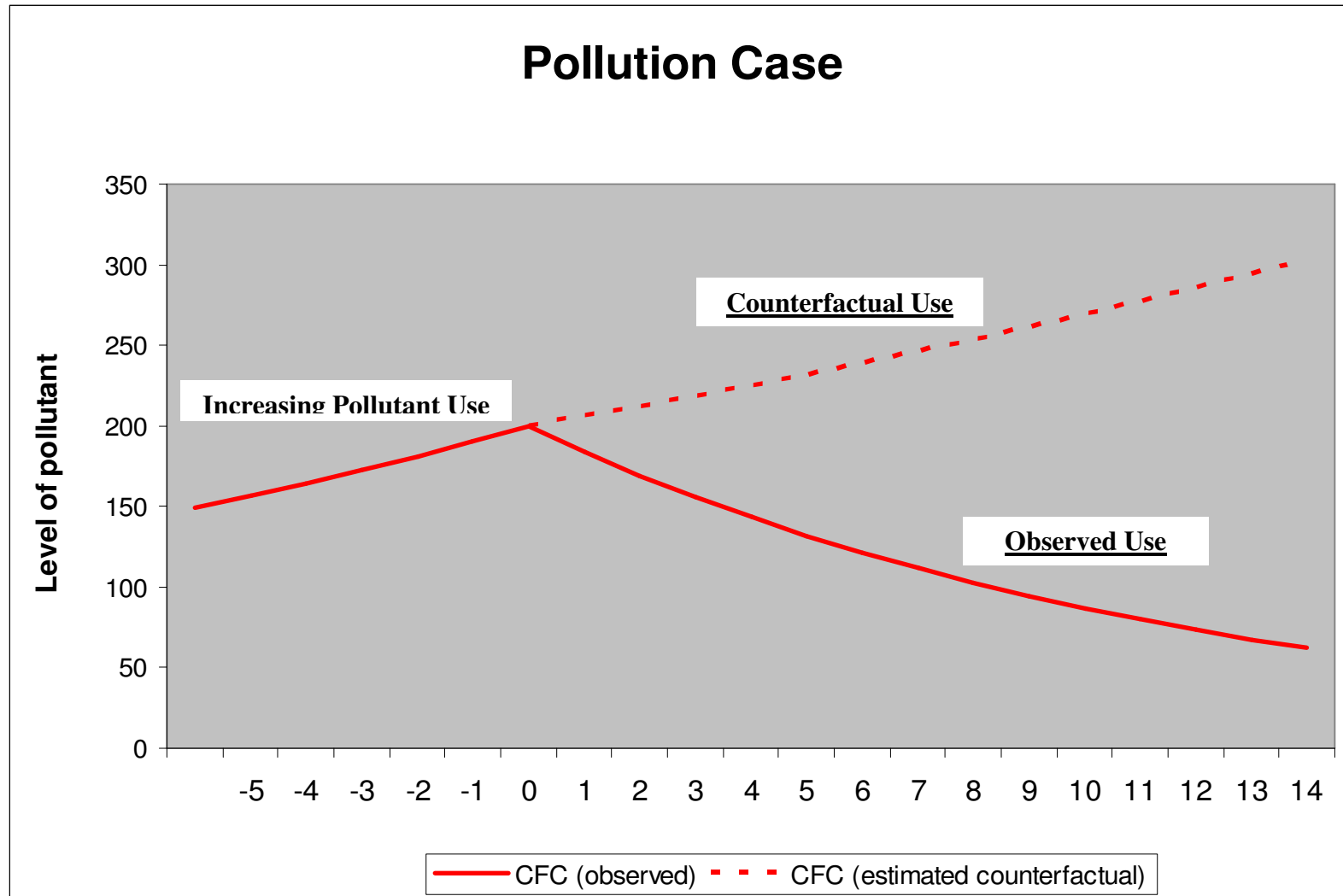
The second implication is more substantive in that it implies that many wildlife agreements sow the seeds of their own failure. This is likely to be particularly true in cases in which economic exigencies dictate that the agreement cannot ban harvest outright (or limit it to particular contexts, such as indigenous cultures or scientific purposes) but must allow commercial harvest but limit its level. Outright bans remove the legal market for a good -- although illegal markets are almost certain to develop, those markets are likely to be dramatically smaller than preexisting legal markets due to both the higher costs that those selling the product are likely to require for the added risk, the greater difficulty that sellers have in identifying buyers, and the normative aversion many buyers will have to engaging in actions that unambiguously violate international law. By contrast, these factors have far less, if any, force for regimes that limit the supply of particular species to legal markets.

### **Tentative conclusions**

This paper has identified some initial obstacles to efforts to evaluate the effects of wildlife agreements. The trajectory of harvest statistics for a hypothetically "fully successful" agreement make generating a counterfactual harvest against which observed harvest can be compared far more difficult than in comparable environmental realms. With pollution regimes, differences in the rate at which emission levels are changing from year to year (i.e., the second

derivative) can be used to compare either members to nonmembers or pre- and post-treaty behavior, and with relative ease can be translated into plausible arguments about whether the agreement had any effect on behavior or not. By contrast, wildlife agreements generally are complicated by two important factors. The first is simply that the trajectories of treaty success and treaty failure are often remarkably similar: a treaty that fails to induce restraint by harvesters will allow the continuation of overexploitation, continuing decline in population, and hence continuing and perhaps quickening declines in harvests; a treaty that succeeds in inducing restraint by harvesters will be evident in similar quickening declines in harvests. In this theoretical landscape, deciphering whether a dramatic decline in harvests in a given realm is due to "mutual restraint, mutually agreed upon" (Hardin 1968) or simply the near extinction of the relevant stock or species becomes a challenging task. Some techniques exist that may begin to allow us to untangle these analytic difficulties. But considerably more effort and thought must be dedicated to such efforts if we are to make significant headway in addressing the precipitous declines visible in almost all wildlife currently being exploited by humans.

**Figure 1**

**Figure 2**

## References

- Biersteker, Thomas. 1993. "Constructing historical counterfactuals to assess the consequences of international regimes: the global debt regime and the course of the debt crisis of the 1980s." In *Regime theory and international relations*, ed. Volker Rittberger. New York: Oxford University Press, 315-338.
- Brown Weiss, Edith, and Harold K. Jacobson, eds. 1998. *Engaging countries: strengthening compliance with international environmental accords*. Cambridge, MA: MIT Press.
- Clapp, Jennifer. 1997. "The illegal CFC trade: an unexpected wrinkle in the ozone protection regime." *International Environmental Affairs* 9:4 (Fall), 259-273.
- Fearon, James D. 1991. "Counterfactuals and hypothesis testing in political science." *World Politics* 43:2 (January), 169-195.
- George, Alexander, and Andrew Bennett. forthcoming. *Case Studies and Theory Development*.
- Hardin, Garrett. 1968. "The tragedy of the commons." *Science* 162:3859 (13 December), 1243-1248.
- Hovi, Jon, Detlef F. Sprinz, and Arild Underdal. 2003. "The Oslo-Potsdam solution to measuring regime effectiveness: critique, response, and the road ahead." *Global Environmental Politics* 3:3 (August), 74-96.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing social inquiry: scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- Miles, Edward L., Arild Underdal, Steinar Andresen, Jørgen Wettestad, Jon Birger Skjærseth, and Elaine M. Carlin, eds. 2002. *Environmental regime effectiveness: confronting theory with evidence*. Cambridge, MA: MIT Press.
- Mitchell, Ronald B. 2002. "A quantitative approach to evaluating international environmental regimes." *Global Environmental Politics* 2:4 (November), 58-83.
- Mitchell, Ronald B. forthcoming. "Quantitative analysis in international environmental politics: toward a theory of relative effectiveness." In *Regime consequences*, eds. Arild Underdal, and Oran Young. London: Kluwer.
- Mitchell, Ronald B., and Thomas Bernauer. 1998. "Empirical research on international environmental policy: designing qualitative case studies." *Journal of Environment and Development* 7:1 (March), 4-31.
- Myers, Ransom A., and Boris Worm. 2003. "Rapid worldwide depletion of predatory fish communities." *Nature* 423:6937 (15 May), 280-283.
- Pauly, Daniel, Villy Christensen, Johanne Dalsgaard, Rainer Froese, and Francisco Torres Jr. 1998. "Fishing down marine food webs." *Science* 279 (6 February), 860-863.
- Sprinz, Detlef, and Carsten Helm. 1999. "The effect of global environmental regimes: a measurement concept." *International Political Science Review* 20:4 (October), 359-369.
- Underdal, Arild, and Oran Young, eds. forthcoming. *Regime consequences*. London: Kluwer.
- Victor, David G., Kal Raustiala, and Eugene B. Skolnikoff, eds. 1998. *The implementation and effectiveness of international environmental commitments*. Cambridge, MA: MIT Press.
- Young, Oran R. 2003. "Determining regime effectiveness: a commentary on the Oslo-Potsdam solution." *Global Environmental Politics* 3:3 (August), 97-104.