Empirical-Quantitative Approaches to the Study of International Environmental Policy

by

Detlef F. Sprinz

PIK - Potsdam Institute for Climate Impact Research
P.O. Box 60 12 03
14412 Potsdam
Germany

Phone: +49 (331) 288-2555/-2532
Fax: +49 (331) 288-2600
E-mail: dsprinz@pik-potsdam.de
URL: http://www.pik-potsdam.de

prepared for

Stuart S. Nagel (ed.):

Global Public Policy: Among and Within Nations

formatting revised, 04 June 1998

Table of Contents

Acknowledgments

An earlier version of this article was presented at the 1997 Annual Convention of the International Studies Association, 18-22 March 1997, The Westin Harbour Castle Hotel, Toronto, Ontario, Canada. The author is indebted to Manfred Binder, Harold Jacobson, and Arild Underdal for comments on an earlier draft.

Abstract

This article provides a concise overview of empirical-quantitative research in the field of international environmental policy. In particular, an ideal research cycle is developed which summarizes the major decisions to be taken in quantitative research. Subsequently, the contemporary quantitative research on international environmental policy and its practical challenges are assessed. The article concludes by emphasizing the various strengths of empirical-quantitative research and providing an outlook on the future of empirical-quantitative studies in the field of international environmental policy.

> *For understandable reasons, case selection in most studies [on international regimes] has been driven by practical considerations instead of methodological requirements. Moreover, the choice of both dependent and independent variables for systematic attention in these small-n case studies has failed, in general, to produce a cumulative and consistent set of information on an agreed upon set of important variables. Each study, in practice, has tended to select idiosyncratic variables, or operationalize common ones in radically different ways. As a result of these limitations, the study of international regimes stands out as somewhat peculiar in its absence of systematic, large-n studies making use of quantitative methods, methods which have advanced the state of the art in almost all other areas of political science.*
>
> (Breitmeier, Levy, Young & Zürn, 1996b, 1)

Why Empirical-Quantitative Research?

Studies of international environmental policy have largely been dominated by empirical-*qualitative* studies, most of which employ the case study design. As Breitmeier et al. (1996b, 1) suggest for the study of international regimes, empirical-quantitative studies may assist in advancing social science research. However, very few empirical-quantitative studies in the field of international environmental policy have been undertaken so far. As Kern and Bratzel summarize

Quantitative methods are the exception in international comparative environmental policy research. ... Most studies are undertaken as case studies or comparative case studies" (Kern & Bratzel, 1996, 53, translation by the author).

As the researchers using empirical-qualitative methods suggest (e.g., Bernauer & Mitchell, 1997), well-crafted case studies require particular care in order to permit valid conclusions. While case studies are likely to generate very detailed information, the degree of confidence in the findings is limited by the small number of cases ("N") which they normally encompass. By contrast, empirical-quantitative research often shows the following characteristics:

(i) medium to large number of cases (from hundreds to thousands),

(ii) multiple explanatory (or exogenous) variables, and

(iii) testability of elaborate structures of hypotheses.[i]

Since experimental settings in international environmental policy hardly exist, quantitative studies will use quasi-experimental designs (Cook & Campbell, 1979).[ii] Empirical-quantitative methods more easily allow the use of "most different case" as opposed to "most similar case" research designs (Przeworski & Teune, 1970), but they have not been necessarily the method of choice: They require explicit theory and data - both of which are not yet abundant in the field of international environmental policy. Conversely, this partially explains the attractiveness of the case study method. Fortunately, even at this point in time, there a few studies in the field of international environmental policy which apply statistical methods, and given increased efforts of data generation, empirical-quantitative research is likely to become more prominent in the foreseeable future. This trajectory may potentially resemble the trend in research on the study of civil and international war as pioneered by the datasets created in the field of international security (e.g., Small & Singer, 1982).

In the following Section, I will summarize the steps to be ideally taken in empirical-quantitative research, and, subsequently, provide a brief overview of empirical-quantitative studies in the field of international environmental policy. Furthermore, the practical challenges these studies face are reviewed. The article concludes with a summary of the advantages of quantitative research and its likely future trajectory.

Designing Empirical-Quantitative Research

Designing an empirical-quantitative study may easily be more demanding than subsequently executing it. The reason for this counter-intuitive conclusion lies in the risk a researcher accepts once s/he has settled for a research design and collected the data. Once data are collected, it may prove difficult to correct for problems of poor model specification, omitted variables, low data quality, missing data based on incomplete responses, or lack of multiple measurements of the same theoretical construct. Therefore, the *design* of an empirical-quantitative study is a crucial step warranting particular attention. Ideally, a researcher should go through the steps outlined in Table 1 in order to strive for high quality results.

Table 1: **The Ideal Sequence of Empirical-Quantitative Research**

Steps
1. Selection of the Phenomenon to Be Explained
2. Development of Theory and Specification of Hypotheses
3. Determination of Data Sources and Scaling Procedures
4. Selection of the Statistical Method for Analysis and Interpretation of Findings
5. Threats to Validity
6. Implications of the Study for Future Research

An overview of each of the steps involved in empirical-quantitative research is provided in the sections to follow.

SELECTION OF THE PHENOMENON TO BE EXPLAINED

The perhaps most important task of any research is to limit oneself to a phenomenon which shall be explained. Alternatively, it may be coined the "dependent variable problem." Much of the research in the field of international environmental policy resolves around international regime "initiation", "implementation" of and "compliance" with the provisions of international environmental agreements at the domestic level, as well as the "dynamics" and "effectiveness" of international environmental regimes. Once the researcher has settled for a phenomenon to be explained, it is advisable to compose a conceptual (as opposed to operational) definition of the dependent variable and explore the various dimensions of the phenomenon under research.

Two principal challenges are associated with selection of the scope of the analysis: A project becomes too broad or too narrow. The first problem practically leads to the challenge of actually executing a very encompassing analysis for reasons of data requirements, costs of research, and demands on theoretical guidance. The second problem of settling for an overly narrow research phenomenon may lead to the lack of interest by the scholarly or policy community due to its high degree of specificity. Ideally, any research should make a contribution to its more narrow confines as well as the advancement of the frontiers of the discipline more broadly. Settling for the appropriate scope of a research project is a delicate balancing act involving consideration of the advancement of theory, methodology, usefulness for public policy, executability, and ability to secure financial support.

DEVELOPMENT OF THEORY AND SPECIFICATION OF HYPOTHESES

After carefully delimiting the phenomenon under investigation, particular care has to be invested in explaining under which circumstances or conditions particular phenomena are expected to occur and under which circumstances are expected *not* to occur. The derivation of a set of hypotheses, ideally amounting to a theory, as well as the derivation of necessary and sufficient conditions (see below) constitute major steps in the research process.

Developing a clear set of hypotheses requires two major components, namely the specification of the exogenous (or predictor) variables as well as the expected directional relationship between the exogenous and the endogenous variable.

Quite often, this amounts to writing an equation (for subsequent statistical estimation). The most generic equation is

$$Y = \alpha + \sum_{i=1}^{n} \beta_i X_i + \varepsilon,$$

where Y is the dependent variable, α the absolute term (i.e., the estimated value of Y if all explanatory variables X_i have assumed default values), X_i represents the particular exogenous or independent variable *i* (range of i = 1, ..., i, j, ..., n), β_i is the coefficient for the explanatory variable X_i, and ε represents the error term. In a dynamic context, time subscripts would be added.[iii]

Specification of the independent variables rests both on the state of the art in the discipline (Do well-explored relationships already exist?) as well as the novel aspects to be added by the particular research proposed. Therefore, finding predictor or explanatory variables is both an act of reviewing the literature as well as a process of intellectual innovation, the latter focusing on hitherto omitted variables, badly operationalized variables, or the development of more elaborate (structural) models (see below). A body of compatible hypotheses can amount to a theory as long as this set of hypotheses is complementary rather than rival. If the research is geared to testing rival hypotheses, then part or all of the relationships hypothesized will have different signs for specific β_i's *or* suggest the non-refutation of a null hypothesis (most often: β_i=0). Including both complementary and rival components in a research design may assist in clarifying to which degree theoretical developments will muster the challenge of empirical-quantitative research. Under ideal circumstances, "conclusive" tests allow for the settling of theoretical disputes in case of rival explanations.

As Bernauer and Mitchell (1997) suggest, purely bivariate relationships are often the point of departure for data analysis, but more elaborate models involve a complex structure of relationships. Introducing intervening and interacting variables provide a transition to what is often called structural modeling.

Intervening variables (IV) are positioned "in between" an exogenous (X_i) variable and an endogenous variable (Y) (see Figure 1). Intervening variables enhance the plausibility of theories by permitting a more detailed structure of relationships between X_i and Y.

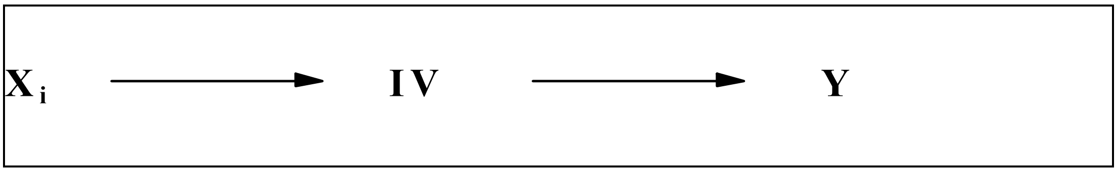

Figure 1: **Intervening Variable (IV)**

By contrast, *interacting variables* (IA) focus on the direct relationship between X_i and Y, however the strength and sign of this relationship between X_i and Y may be contingent on the level of the IA variable (see Figure 2). Depending on the level of the IA, the measure of covariation between X_i and Y may vary substantially, and the sign of the relationship may even change across levels of IA.

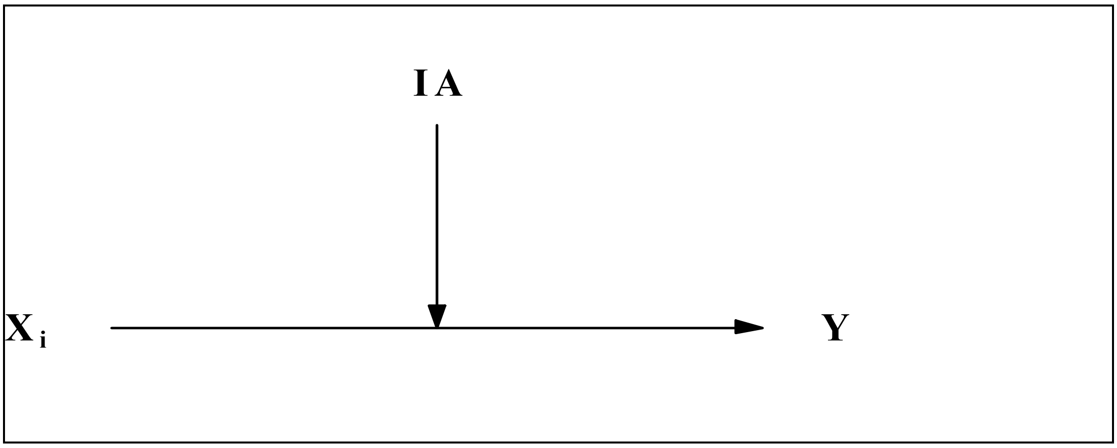

Figure 2: **Interacting Variable (IA)**

Incorporating intervening and interacting variables leads the researcher to contemplate more complicate model structures, ultimately leading to so-called "structural equation models", often referred to as path analysis or, in more elaborate fashion, structural equation models. These combine measurement components for each variable complex (e.g., the various variables representing environmental problem pressure) with a structure of relationships between the variable complexes, in particular among the set of independent, intervening, and dependent variables.

By reflecting about hypotheses, the researcher has to contemplate if the independent variables constitute necessary or sufficient conditions for the dependent variable. Whereas necessary conditions focus on the occurrences of the dependent variable, sufficient conditions focus on the occurrence of the independent variable.[iv] In empirical research, we are unlikely to find that some independent variables are simultaneously necessary and sufficient conditions for a particular outcome, however, by clearly thinking about necessary and sufficient conditions at the stage of research design, the researcher is able to build expectations about the strength of association among variables -- which ultimately supports interpretation of findings at a later stage.

Overall, the stage of theory development and derivation of hypotheses clarifies the scope of the analysis to be undertaken. Together with the previous step of problem selection, subsequent data generation and evaluation can now be tackled. Regardless of the empirical method chosen, the first two steps of the ideal research cycle are necessary and sufficient conditions for conducting fruitful empirical analyses and ideally contribute to a cumulative research program.

DETERMINATION OF DATA SOURCES AND SCALING PROCEDURES

Empirical-quantitative methods normally rely on large datasets (large "N" studies) in contrast to the case study method. These data refer to "units of analysis," i.e., the underlying unit (such as countries, organizations, or geographical grid) for which particular variables are collected. The unit of analysis is also related to the "level of analysis" (Singer, 1969) as explanations can be potentially nested across levels of analysis (e.g., Bryk & Raudenbusch, 1992).

While the previous two steps have led the researcher to specify the variable constructs to be included in the analysis, they have to be operationalized at this step of research. This involves specification of which operational variables actually represent particular concepts, determination of the scaling of such variables, checking for the reliability and validity of such data, and the evaluation of existing data collections and determination of additional data needs (if any).

Operationalizing concepts necessitates attention being placed on the reliability and validity of the measures.

> Fundamentally, *reliability* concerns the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials. The measurement of any phenomenon always contains a certain element of chance error. The goal of error-free measurement - while laudable - is never attained in any area of scientific investigation (Carmines & Zeller, 1986, 11, emphasis in the original).

In my own research, for example, I was surprised to encounter difficulties in reliable measures of the Gross Domestic Product (GDP) for industrialized countries - despite more than 40 years of experience with this concept (see also Binder, 1996b, 156-160). These reliability problems are likely to be even more pronounced in fields with substantially shorter traditions in operationalizing variables such as "regime effectiveness" or "environmental problem pressure."

While measures may be reliably measured, they may not be valid measures of the underlying concept. As Carmines and Zeller define:

> In a very general sense, any measuring device is valid if it does what it is intended to do. An indicator of some abstract concept is *valid* to the extent that it measures what it purports to measure (Carmines & Zeller, 1986, 12, emphasis added by the author).

The Gross Domestic Product regrettably measures only market transactions and omits, inter alias, non-market transactions, depletion of natural resources, or changes in environmental quality. Thus, its validity as a measure of economic wealth (as a yearly flow) is undermined. Because direct measurements (e.g., flow of water at a particular physical location) are rare in social science research, we will often have to resort to imperfect measurement - which poses questions of external validity, esp. construct validity (see below).[v]

Besides considerations for reliability and validity, variables normally adhere to some underlying scoring scheme. Nominal, ordinal, interval, or cardinal scales have to be distinguished, and the variables may be either uni- or multidimensional.

Nominal scales do not espouse a particular order (e.g., Christian, Hindu, Jewish, or Muslim orientation), whereas the other scales pose increasing demands on the variables. While *ordinal* measures only require the ordering of variables (e.g., in ascending or descending order) and intervals between scores are not fixed, the latter are fixed for *interval* scales; *cardinal* scales have a "natural" zero point of reference (such as 0 degree Kelvin or 0 US Dollar) in addition to the properties of interval scales. While the most powerful statistical techniques ideally require interval scale variables, data often do not show these properties. Instead, the specific technique for data analysis has to be adjusted to the type of existing data.

Besides determining the scale, variables (constructs) may be of single or multiple dimensions. The best known *unidimensional scales* are Likert scales (e.g., a five point scale of the degree of approval for environmental legislation) or Guttman scores (a cumulative score representing the highest score on the underlying dimension). *Multidimensional scales* (e.g., representing environmental problem pressure for a specific region) essentially refer to factor-analytic methods which deduce underlying dimensions from a larger set of variables. In practical terms, most studies of international environmental policy are unidimensional, and if they are multidimensional, they are rarely constructed as such explicitly.[vi]

Before actually turning to data collection activities, the *sampling* strategy has to be considered. With a perspective towards generalization, a random sample of the relevant population allows for the strongest degree of statistical conclusion validity (see below) and inferential statistics essentially work on samples - not the population. In practice, data limitations often preclude the derivation of satisfactory random samples, and the interpretation of the results should take this caveat into account. Quite often, the particular research design will require additional data collection efforts beyond those hitherto undertaken. As with existing data sources, researches have to attend to the various aspects of reliability, validity, scaling, dimensionality, and sampling strategy referred to in this section, and pretests are highly recommended.

SELECTION OF THE STATISTICAL METHOD FOR ANALYSIS AND INTERPRETATION OF FINDINGS

Upon finalizing data selection and collection, the researcher can actually turn to analyzing the data. While descriptive (univariate) statistics (such as standard deviation or variance, skewedness, and kurtosis) are quite informative for a first exploration of the data, theories can only be tested with inferential statistics which relate independent to dependent variables. A variety of statistical methods are at the researcher's disposal, and the interested reader is referred to the specialized literature.[vii]

The central aim is to find coefficients for the structural relationships between independent and dependent variables -- without violating central assumptions of the statistical technique chosen. Most prominent are variants of the General Linear Model with Ordinary Least Square (OLS) estimators or, when important assumptions of OLS are violated, the use of Maximum Likelihood (ML) estimators. Besides linear relationships among independent and dependent variables, the analysis of qualitative and limited (score) dependent variables requires special techniques, including contingency table analysis, logit and probit models, as well as logistic regression (see, e.g., Fienberg, 1980; Hosmer & Lemeshow, 1989).[viii]

What appears to be most important in the context of research on international environmental policy is the difference between static and dynamic forms of analysis, with the former concentrating on one particular point in time whereas the latter including observations at multiple points in time. For lack of appropriate data, essentially dynamic processes are often tested with the help of crossectional data (see below).

Crucial to any preliminary analysis is to undertake a sensitivity analysis, i.e., an investigation of the set of resulting parameters and confidence intervals if some crucial aspect of the tested model or the particular statistical method chosen is changed. Only if the results are insensitive to such changes, then the researcher may have heightened confidence in her or his statistical findings. Following the statistical analysis of data, the results need to be interpreted. This consists of two important parts, namely the substantive interpretation and the statistical interpretation. In this Section, I will cover the latter aspect in this section and defer the former until the following Section.

In the statistical interpretation of findings, a range of criteria are used. Most techniques of inferential statistics are geared to maximize the explained part of the variance found in the dependent variable or minimizing the discrepancy between the matrix of observed and estimated relationships. Most often, statistical criteria, such as percentage of variation of the dependent variable explained (so-called corrected R^2), the mean squared error of the equation, and significance tests of the parameters are employed. However, also the sign and the magnitude of effects of the independent variables play an important role, esp. if they can be easily expressed as changes on the scale of the dependent variable (i.e., unstandardized scores).

Particular attention is regularly paid to significance tests of the parameter estimate of independent variables or the overall equation (F-test). In particular, for larger N studies, the 5% probability of erroneously refuting the null hypothesis (of no effect) is accepted as a statistical criterion -- with even smaller error probabilities preferred. While two-tailed tests are more demanding in statistical terms, one-tailed test - based on theoretical reasoning (see above) - are equally acceptable. As the number of cases (N) is increasing, it is commonplace to find higher significance levels for purely statistical reasons. However, large N studies are not very common in the research on international environmental policy to date which lets us expect that significance levels are comparatively low.

Most software packages include standard diagnostics which check for classical estimation problems such as non-normal distribution of the error term ε, multicollinearity (i.e., high correlation among independent variables), heteroskedasticity (i.e., lack of a constant average error term), autoregressive effects over time between the dependent and independent variables, etc. (for details see, e.g., Kennedy, 1992). Not included in statistical software packages is the experience which a researcher brings to the substantive and data domain. One of the most useful exploratory techniques in data analysis remain scatterplots of the distribution of each variable, the bivariate plots between (i) independent and dependent variables as well as (ii) between independent variables. These techniques have been advanced by contemporary exploratory data analysis software for three-dimensional perspectives and offer an efficient way to explore potential data and estimation problems in the beginning of a comprehensive statistical analysis.

THREATS TO VALIDITY

In evaluating the validity of the statistical findings, the researcher makes important decisions about the implications of her or his research in terms of theory. Cook and Campbell (1979) differentiate four different kinds of validity for experimental and quasi-experimental research:

> *[S]tatistical conclusion validity* refers to inferences about whether it is reasonable to presume covariation given the specified α [the chosen probability thresholds for rejecting the null

hypothesis] and the obtained variances (Cook & Campbell, 1979, 41, emphasis in the original, additions by the author).

[I]nternal validity refer[s] to the validity with which statements can be made about whether there is a causal relationship from one variable to another in the form in which the variables were manipulated or measured (Cook & Campbell, 1979, 38, emphasis in the original).

[C]onstruct validity of causes or effects ... refer to the approximate validity with which we can make generalizations about higher-order constructs from research operations (Cook & Campbell, 1979, 38, emphasis in the original).

[E]xternal validity ... refer[s] to the approximate validity with which conclusions are drawn about the generalizability of a causal relationship to and across populations of persons, settings, and times (Cook & Campbell, 1979, 39, emphasis in the original).

Threats to validity take a range of forms, and I will cover only a few aspects which seem to apply to the quantitative studies of international environmental policy in particular.[ix]

Threats to *statistical conclusion validity* concern the statistical power of tests, as well as the magnitude of relationships between independent and dependent variables. These threats do arise if the number of cases (N) is low, assumptions of statistical tests are violated (see above), and unreliable measures are chosen.

Threats to *internal validity* target the causal order of relationships, including the status of intervening and interacting variables (see above). Omitting important but unobserved variables, problems with bottom and ceiling effects (e.g., there are logical minima [0%] and maxima {100%] for connecting a population to piped drinking water), and selection effects (e.g., signatories to international environmental agreements are more likely to adhere to those provisions than non-signatory countries) are cases in point.

By contrast to the previous two types of validity, *construct validity* refers to the problem of determining if operational measures actually measure the concept they are supposed to represent. This applies equally to the dependent as well as independent variables. Particular threats to construct validity arise if only one operationalization per construct is used in the analysis.

Finally, *external validity* refers to the generalizability of findings across larger domains of units of analysis, time periods, and geographical settings. However, sweeping generalizations are often hindered by interaction between selection and treatment (e.g., selecting wealthy OECD countries and focusing on the rate of connection to residential sewerage systems) and interaction of history and treatment (e.g., the effect of the Chernobyl disaster on national nuclear policies for recipient countries of radiation as compared to those countries which did not receive unwanted radiation from Chernobyl). Random sampling of cases across the universe or introducing heterogeneous cases into the analysis are ways to increase external validity.

Theoretical clarity at the beginning of the research process helps limit the impact of the various threats to validity, however, it is ultimately the researcher and his or her critics who will decide to which degree the quantitative analysis has been a proper test of a specific theory.

IMPLICATIONS OF THE STUDY FOR FUTURE RESEARCH

While any quantitative research will necessarily have to pass through the previous five steps at least once, in reality, the research process is more like a recurring cycle. In a cumulative research program which allows for replications and expansion of prior analyses, researchers gain an increased familiarity with the theoretical domain, the data domain, and the feasibility of research at a particular point in time. Actively considering threats to validity is an important source for improving future research and to contemplate more appropriate tests of the theory, extending its geographical and temporal domains, and probing the applicability of a theory in another substantive domain. In this sense, the research cycle is a learning process for researchers and subsequent generations of researchers. As compared to empirical-qualitative methods, it is worth mentioning that quantitative methods enjoy much stricter data definitions, provide documented datasets which allow replication with the same and different statistical methods, and permit replication of the research findings with equivalent alternative datasets, as well as sensitivity analyses.

Empirical-Quantitative Research On International Environmental Policy

In comparison to the state of research on international and civil war as well as international economic relations, the study of international environmental policy is still at an early stage in applying quantitative methods, simulation studies, or game-theoretic approaches. To this date, the field is dominated by empirical-qualitative studies employing the (comparative) case study method (see Bernauer & Mitchell, 1997). However, since the early 1990s, three groups of empirical-quantitative studies have emerged, namely

(i) comparative national environmental policy studies and national implementation studies,
(ii) domestic-international linkages, and the
(iii) international-domestic perspective.

Falling into the first category, the team led by Jänicke at the Research Unit for Environmental Policy at the Free University of Berlin (Germany) has undertaken a series of studies on the effect of increasing wealth (measured as Gross Domestic Product) as well as the effects of changed industry structures on the environment. In particular, this group of researchers has focused on industrialized countries (OECD countries as well as Central and Eastern Europe) and probes the question if increased wealth has led to reduced or increased environmental pressures (Jänicke, 1996; Jänicke et al., 1993; Jänicke, Mönch, Ranneberg & Simonis, 1988). Their composite hypothesis essentially states that as poor countries become richer, they first increase environmental pressures until a turning point in their economic development; beyond this point, countries reduce environmental pressures as they become even richer. We may wish to term this hypothesis the "environmental stages of economic development." Jänicke and collaborators conclude that for some indictors, increases in wealth have led to improved environmental performance, while in other cases increasing wealth has led to a worsening environmental performance. Furthermore, in few cases, the "environmental stages of economic development" hypothesis seems to hold (Jänicke, Mönch & Binder, 1996, 129; Jänicke et al., 1993, 48-49).[x]

By contrast to the comparative studies by Jänicke and collaborators, Widmer (1991) uses a time series approach in order to investigate if regulations of sulfur emissions in Switzerland are actually implemented and if air quality subsequently improves. He finds that, in particular, public policies on air pollution are achieving their goals if they are augmented by strict governmentally set limit values on emissions, whereas implementing agents deviate in case of less strict types of regulations (Widmer, 1991, 184-185). And in a much earlier study, Crandall shows with the help of roll call data that regional interests are clearly reflected in US Congress voting patterns regarding the type of regulations furthered on federal air pollution regulations (Crandall, 1983, ch. 7).

While the previous two studies are purely domestic in scope, a second track of research has focused on the domestic-international link by relating domestic political variables to national decisions on international environmental agreements. In the most abstract form, problem pressure ("ecological vulnerability") and the costs of regulation explain why countries are willing to sign stringent international environmental agreements as opposed to not signing them in the context of the United National Economic Commission for Europe during the second half of the 1980s (Sprinz, 1993; Sprinz & Vaahtoranta, 1994; Sprinz, 1992, ch. 7). Subsequent research by the first author inserted a political process model, based on aggregate expert responses, between those two macro factors. This enabled, on the one hand, environmental problem pressure to translate into public support for international environmental regulation, and increasing costs of regulation into support for not signing international regulations. The logistic regression results show that this operationalization of the various domestic political pressure components explains the regulation of sulfur emissions more convincingly than those for nitrogen oxides (Sprinz, 1992, ch. 6; Sprinz, forthcoming-a).

Subsequent research by Sprinz and Helm in the same case domain was geared towards isolating the effect of international institutions (United Nations Economic Commission for Europe) on national emission reductions goals beyond what countries would have undertaken in the absence of the international regime (Sprinz & Helm, 1996, ch. 2). The systematic counterfactual analysis allows for a structured scoring mechanism for signatories and non-signatories. Further research explored the various monitoring functions of regimes, esp. monitoring of environmental quality and emission reductions, and showed that both domestic and international sources of monitoring are roughly of equal importance for sulfur and nitrogen oxides during the second half of the 1980s and the first half of the 1990s (Sprinz & Helm, 1996, ch. 3).

A third group of studies, led by Jacobson and Brown Weiss, focuses on the degree of national compliance with international environmental agreements (Jacobson & Weiss, 1995; Jacobson & Weiss, 1996; Jacobson & Weiss, forthcoming-a). They selected five international natural resource and pollution agreements for nine political units. By carefully justifying their selection of the nine political units and the five international environmental agreements (Jacobson & Weiss, 1996, 2-4), they were using a "most different case" design for both the political units as well as the substantive domains chosen. As a consequence, we should expect external validity (within the domain of international environmental policy) to be quite high. While the project has not yet published its detailed statistical (rather than the substantive) findings, its datasets also incorporate a temporal domain - thereby extending the number of cases in comparison to present studies of international environmental policy to a degree which allows the use of quantitative methodology for crosstemporal analysis.

While these three groups of studies summarized above clearly outline the phenomenon to be explained, the status of theory and clear specification of *ex ante hypotheses* is not always equally well developed. For example, in deriving a measure of the success of environmental policy, Binder suggests a range of factors which are expected to explain the reduction of sulfur dioxide emissions in industrialized countries. The analysis of the variation of the success of environmental policy rests largely on macroeconomic variables (which are not environmental policy variables), timing of institutionalization of environmental policy, research expenses, demographic trend, and public attitudes on air pollution as explanatory variables (Binder, 1996b). While the author only seeks to describe (rather than explain) the effect of these variables on sulfur dioxide emissions related to fossil fuel combustion (ibid.; personal communication, 20-21 March 1997), theory development is normally geared towards explanation. While exploratory analysis is a necessary step at an early stage of research, more advanced analyses require a sufficiently well developed theoretical body undergoing subsequent hypothesis testing. In terms of theory development, the insufficient inclusion of important variables is perhaps the most regrettable. More generally, omitted variables have unfortunate effects on statistical results, esp. on the magnitude and direction of variable coefficients already included into the analysis. A more theory-driven research design would have focused on the explanatory effect of a subset of policy variables, esp. since the lager research project is geared towards explaining successful environmental policy. In particular, including governmental regulations of the specific pollutant under investigation could prove particularly rewarding.

The *data* available in the field of international environmental policy are often of less than ideal quality. Even some of the best explored data collection efforts, such as those on measuring the Gross Domestic Product, display substantive discrepancies across sources. Regularly, researchers in the field have to undertake time-consuming data collection activities themselves, including the International Regimes Database project (Breitmeier, Levy, Young & Zürn, 1996a; Breitmeier et al., 1996b), the Compliance with International Environmental Accords project (Jacobson & Weiss, 1995; 1996; forthcoming-b), or structured expert interviews (Sprinz, 1992, ch. 6). More often than not, the data collections to this very date contain rather small numbers of cases which ultimately hamper the utilization of the full power of statistical methods.

In practice, small numbers have two effects: Significance levels will be low in comparison to large N studies involving hundreds of observation and the statistical method may be prove inappropriate.[xi]

First, in his research on explaining why countries sign international environmental agreements, Sprinz aggregates expert responses to national means which led to the standard problem of too many predictors (5) relative to cases (N=9) in a logistic regression analysis (Sprinz, 1992, ch. 6). While the use of so-called "induced variables"[xii] (Alwin, 1988) helped in reducing the number of predictor variables, important information is necessarily lost by this method, and low significance levels may result in not rejecting a null hypothesis - which may be more easily rejected for a larger set of cases.

Second, more courageous are the exploratory analyses by Binder (1996b, 197-201). For a set of 18 cases with six predictors, he estimates a structural equation model with LISREL in order to account for the success of environmental policy on the country level. The introduction of additional cases would certainly be advisable in view of the statistical method chosen.

Introducing additional cases may actually have diametrically opposite effects: One the one hand, it *may* (but does not have to!) increase the significance levels of parameter estimates, thereby increasing confidence in the theoretical propositions of the model. On the other hand, a previously empirically fitted model may suffer from reduced explanatory power and reduced overall fit if the additional cases introduce variation which had previously been insufficiently incorporated in the analysis. It therefore critically depends on the degree to which the original sample is representative of the true (but unobserved) distribution of properties in the population.

However, sometimes quantitative methods are not extensively used although this would be possible. In their crossnational research on the impact of economic development on indictors of environmental degradation, Jänicke et al. use a "path comparison diagram" which only utilizes data from the beginning and the end of a longer time period (Jänicke et al., 1996; Jänicke et al., 1993, ch. II) -- rather than more appropriate time series techniques, including pooled time series.[xiii] While the "path comparison diagram" (i.e., linking the value of variables at the beginning of a period with the value at the end of a period by way of a linear arrow) assists quick visual inspection (esp. for an exploratory analysis), it does not guard against the effects of outliers at the beginning and the end of the period, suggests linear trends when they are not necessarily present, and uses a crossnational design for a hypothesis which is essentially dynamic. While the decision in favor of the path comparison diagram may be driven by data availability, focusing on equivalent time periods of a country's economic development may yield stronger rather than weaker results in view of the hypotheses put forward. Furthermore, the same visual graphs can be interpreted differently for lack of a standardized method of analysis. For example, for cases where Jänicke et al. suggest that increasing wealth leads to higher levels of pollution, the graphs for some pollutants (e.g., NO_x, volatile organic compounds, and residential waste) (Jänicke et al., 1996) seem to indicate that at least the rate of growth of pollutants is decreasing or the level is decreasing for most countries. Therefore, a more rigorous method for evaluation may assist more systematic and reproducible evaluation of the same data. Furthermore, some of the data may suffer from censoring because if the researchers had longer time series, it appears plausible that some of the benign effects would become more pronounced.[xiv]

The potentially greatest problems arise in conjunction with the *validity of findings*. The examples mentioned further above with regard to the low number of explanatory variables relative to the number of observation as well as unnecessarily simple methods of analysis point to problems with statistical conclusion validity. While problems of influential outliers and their effects on results are often well documented (e.g., Binder, 1996b; Sprinz, 1992, ch. 5-6), this cannot be said of problems of heteroskedasticity (e.g., Binder, 1996b).[xv]

More challenging than statistical conclusion validity is the assessment of internal validity, since the causal status of the empirically observed regularity is to be assessed. According to John Stuart Mills

> causal inference depends on three factors: first, the cause has to precede the effect in time, second, the cause and effect have to be related; and third, other explanations of the cause-effect relationship have to be eliminated. The third criterion is a difficult one (Cook & Campbell, 1979, 18).

Ruling out alternative explanations of the observed statistical association is, in a strict sense, practically impossible or impracticable (see also Asher, 1983, 12). Therefore, the status of theory and the body of empirical studies in the particular field assist in judging the plausibility of the causal structure of arguments.[xvi]

Threats to construct validity may account for much of the disputes within disciplines. What is an "international regime" in operational terms? What is a measure of the "success of environmental policy"? What is a valid measure of a nation's "wealth"? In some research seminars on international conflict, the proverb is rumored: "You live by your coding rules, and you die by your coding rules" (Anonymous).

Given such implications for construct validity, clarity and plausibility of variable definitions matter. While I have already referred to the problems of using the Gross Domestic Product as a yearly flow measure of national wealth, I will briefly discuss the other two examples, namely the definition of international regimes and the operationalization of successful environmental policy.

According to the most widely accepted definition,

> [international] regimes can be defined as sets of implicit or explicit principles, norms, rules, and decision-making procedures around which actors' expectations converge in a given area of international relations. Principles are beliefs of fact, causation, and rectitude. Norms are standards of behavior defined in terms of rights and obligations. Rules are specific prescriptions or proscriptions for actions. Decision-making procedures are prevailing practices for making and implementing collective choice (Krasner, 1983, 2).

While this theoretical definition has proven useful for a large research program on international regimes, the specific criteria for inclusion and exclusion under the term "international regime" have remained vague. While Breitmeier et al. elaborate this critical point by asking "What is a case?" (Breitmeier et al., 1996b, 4-9), they avoid providing a *precise* answer. This leads to unreliable measurement of the unit of analysis and leaves the researcher with considerable discretion with regard to case inclusion.

Conversely, it may be even more challenging to operationalize what successful environmental policy is. Ideally, specific political or other interventions should ultimately yield an improvement in the state of the environment (environmental quality). Empirically, it is often more convenient to resort to an analysis of pollution data - with the implicit understanding that less pollution loads normally result in an improved environmental quality in the long run. In his effort to develop an indicator of successful environmental policy, Binder justifies the selection of sulfur dioxide emissions resulting from fossil fuel combustion on three grounds:

(i) well documented and reliable data,

(ii) influence of environmental policy can be easily identified, and

(iii) in comparison to other pollutants, the decline in sulfur dioxide emission has been the strongest (Binder, 1996a, 143-144).

While the existence of high quality and complete data is very desirable, it remains unclear why successful environmental policy has to be limited to one pollutant only. However, even if this were accepted by tilting the research question towards the explanation why sulfur dioxide emission reductions occurred in industrialized countries, it is striking to observe that the studies do not use explicit environmental policy variables to explain emission reductions (see Binder, 1996a; Binder, 1996b). In conclusion, construct validity for "successful environmental policy" may be highly limited because it restricts the pollution domains to a minimum, potentially confuses independent and dependent variables, and largely omits prima face environmental policy variables from the explanation.

In general, threats to construct validity also impede external validity, i.e., inferences across larger substantive domains of research, geographical domains, and time periods. For example, in order to make the measurement of regime effectiveness (Sprinz & Helm, 1996) more useful to other domains of international political economy, the equivalency of the research protocol for these domains needs to be explored.

The Future of Empirical-Quantitative Research on International Environmental Policy

High quality empirical research is demanding - regardless of the choice being made between qualitative and quantitative methods. In practical terms, quantitative research requires researchers to make a range of explicit decisions which constitute comparative "advantages" of this method (see Table 2).

First, quantitative methods force researchers to think very clearly about what they wish to explain, direction of hypotheses, and how this forms part of complementary or rival theories.

However, there is also a potential danger involved by researchers being tempted by exploratory data software or options of automatic model specification (in structural terms) to arrive at ex post theoretical models. Hypothesis testing requires, strictly speaking, an ex ante theory.

Table 2: **The Advantages of Empirical-Quantitative Methods**

Advantages
clear specification of theoretical model (ex ante)
specification of units of analysis, temporal domain, and geographical domain
operationalization of theoretical concepts
clearly defined methods of (statistical) inference; potential for replication and sensitivity analysis
visualization of data and results

Second, while specification of the units of analysis as well as the temporal and geographical domain should be undertaken irrespective of the particular empirical research method chosen, construction of a data base requires clear delimitations.

Third, quantitative methods ideally induce researchers to clearly define the operational form of theoretical constructs. While this often involves compromises in view of data availability, time, and resource requirements, it provides a bridge between the theoretical and empirical sphere. Conversely, this step is a potential source of serious disagreement among researchers regarding the construct validity of a measure.[xvii]

Fourth, statistical methods limit arbitrariness in inference as long as the appropriate statistical method is chosen and crucial assumptions of a method are not violated. Only under very limited circumstances will qualitative methods be equally successful in the domain of (statistical) inference.

Fifth, with the advance of graphical methods of data presentation and exploration, it becomes be easier to communicate findings to non-expert audiences.

Sixth, it may be worth remembering that given the state of research on international environmental policy, far reaching generalizations may be tempting - but rarely warranted.

Up to now, data for industrialized countries dominate analyses with a focus on the post-World War II period; the projects led by Breitmeier et al. (1996a; 1996b), Jacobson and Brown Weiss (Jacobson & Weiss, 1995; Jacobson & Weiss, 1996; Jacobson & Weiss, forthcoming-b), and Jänicke (Jänicke, 1996; Jänicke et al., 1993) are likely to expand our data domains substantially. International governmental and non-governmental organizations (such as the Consortium for International Earth Science Information Network (CIESIN), European Union, OECD, UNEP, World Bank, and World Resources Institute) collect various important economic, political, social, and ecosystem variables. Among these sources, the CIESIN databases focus most explicitly on international environmental agreements.[xviii] Finally, more relevant data are collected as a sideeffect of research on sustainable development (Hammond, Adriaanse, Rodenburg, Bryant & Woodward, 1995; UNDP, 1994; World Bank/Environmentally Sustainable Development, 1995; World Resources Institute, 1996).

It seems fair to conclude that governments will search for well corroborated research results in order to spend their precious resources more efficiently in the pursuit of international environmental policies. Empirical-quantitative methods, as part of a progressive research program, may make useful contributions both to the state of theory and international public policy in order to more successfully manage Planet Earth.

References

Alwin, D. F. (1988). Measurement and the Interpretation of Effects in Structural Equation Models. In J. S. Long (Ed.), Common Problems/Proper Solutions - Avoiding Error in Quantitative Research, (pp. 15-45). Newbury Park, CA: Sage.

Asher, H. B. (1983). Causal Modeling. (second ed.). Beverly Hills, CA: Sage.

Bartels, L. M., & Brady, H. E. (1993). The State of Quantitative Political Methodology. In A. W. Finifter (Ed.), Political Science: The State of the Discipline II, (pp. 121-159). Washington, D.C.: The American Political Science Association.

Bernauer, T., & Mitchell, R. (1997). Empirical Research on International Environmental Policy: Designing Qualitative Case Studies. Paper Presented at The Westin Harbour Castle Hotel, Toronto, Ontario: International Studies Association.

Binder, M. (1996a). Die Operationalisierung umweltpolitischen Erfolges: Probleme und Lösungsansätze [Operationalizing Successful Environmental Policy: Problems and Partial Solutions]. In M. Jänicke (Ed.), Umweltpolitik der Industrieländer [The Environmental Policy of Industrialized Countries], (pp. 134-152). Berlin: Edition Sigma.

Binder, M. (1996b). Mögliche Erfolgsbedingungen der Schwefeldioxidminderung im internationalen Vergleich [An International Comparison of Potential Prerequisites of Successful Reductions of Sulfurdioxide Emissions]. In M. Jänicke (Ed.), Umweltpolitik der Industrieländer [The Environmental Policy of Industrialized Countries], (pp. 153-201). Berlin: Edition Sigma.

Breitmeier, H., Levy, M. A., Young, O. R., & Zürn, M. (1996a). International Regimes Database (IRD): Data Protocol (IIASA Working Paper WP-96-154). Laxenburg, Austria: International Institute for Applied Systems Analysis.

Breitmeier, H., Levy, M. A., Young, O. R., & Zürn, M. (1996b). The International Regimes Database as a Tool for the Study of International Cooperation (IIASA Working Paper WP-96-160). Laxenburg, Austria: International Institute for Applied Systems Analysis.

Bryk, A., & Raudenbusch, S. W. (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park: Sage.

Carmines, E. G., & Zeller, R. A. (1986). Reliability and Validity Assessment. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-017). Beverly Hills, CA: Sage Publications.

Cook, T. D., & Campbell, D. T. (1979). Quasi-Experimentation: Design & Analysis Issues for Field Settings. Boston, MA: Houghton Mifflin Company.

Crandall, R. W. (1983). Controlling Industrial Pollution - The Economics and Politics of Clean Air. Washington, D.C.: The Brookings Institution.

Darnell, A. C. (1994). A Dictionary of Econometrics. Aldershot: Edward Elgar.

Fienberg, S. E. (1980). The Analysis of Cross-Classified Categorical Data. (second ed.). Cambridge, MA: MIT Press.

Greene, W. H. (1997). Econometrics Analysis. (third ed.). Upper Saddle River, N.J.: Prentice Hall.

Hammond, A., Adriaanse, A., Rodenburg, E., Bryant, D., & Woodward, R. (1995). Environmental Indicators: A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development. Washington, D.C.: World Resources Institute.

Hosmer, D. W., & Lemeshow, S. (1989). Applied Logistic Regression. New York, NY: John Wiley.

Jacobson, H. K., & Weiss, E. B. (1995). Strengthening Compliance with International Environmental Accords: Preliminary Observations from a Collaborative Project. Global Governance, 1, 119-148.

Jacobson, H. K., & Weiss, E. B. (1996). Compliance with International Environmental Accords: Achievement and Strategies, (mimeo.). Villa Rigot, Geneva: Graduate Institute of International Studies.

Jacobson, H. K., & Weiss, E. B. (forthcoming-a). Conclusion: Implementation and Compliance: Assessing the Record of Designing Strategies. In H. K. Jacobson & E. B. Weiss (Eds.), Engaging Countries, Strengthening Compliance with International Environmental Accords. Cambridge, MA: MIT Press.

Jacobson, H. K., & Weiss, E. B. (Eds.). (forthcoming-b). Engaging Countries, Strengthening Compliance with International Environmental Accords. Cambridge, MA: MIT Press.

Jacoby, W. G. (1991). Data Theory and Dimensional Analysis. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-078). Newbury Park, CA: Sage.

Jänicke, M. (Ed.). (1996). Umweltpolitik der Industrieländer [The Environmental Policy of Industrialized Countries]. Berlin: Edition Sigma.

Jänicke, M., Mönch, H., & Binder, M. (1996). Umweltindikatorenprofile im Industrieländervergleich: Wohlstandniveau und Problemstruktur [Comparison of Environmental Indicator Profiles of Industrialized Countries: Wealth and Problem Structure]. In M. Jänicke (Ed.), Umweltpolitik der Industrieländer [The Environmental Policy of Industrialized Countries], (pp. 113-131). Edition Sigma: Berlin.

Jänicke, M., Mönch, H., Binder, M., Carius, A., Foltjanty-Jost, G., Götz, N., Ranneberg, T., & Schneller, M. (1993). Umweltentlastung durch industriellen Strukturwandel? Eine explorative Studie über 32 Industrieländer (1970 bis 1990) [Improvement of the Environment As a Result of the Structural Change of the Economy? An Explorative Study of 32 Industrialized Countries (1970-1990)]. (second, revised ed.). Berlin: Edition Sigma.

Jänicke, M., Mönch, H., Ranneberg, T., & Simonis, U. E. (1988). Structural Change and Environmental Impact (FS II-88-402). Berlin: Wissenschaftszentrum für Sozialforschung (Science Center Berlin).

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1985). The Theory and Practice of Econometrics. (second ed.). New York, N.Y.: John Wiley and Sons.

Kennedy, P. (1992). A Guide to Econometrics. (third ed.). Cambridge, MA: MIT Press.

Kern, K., & Bratzel, S. (1996). Umweltpolitischer Erfolg im internationalen Vergleich: Zum Stand der Forschung [An International Comparison of Research on Successful Environmental Policy: The State of the Art]. In M. Jänicke (Ed.), Umweltpolitik der Industrieländer [The Environmental Policy of Industrialized Countries], (pp. 29-58). Berlin: Edition Sigma.

Kim, J.-O., & Mueller, C. W. (1986a). Factor Analysis - Statistical Methods and Practical Issues. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-014). Beverly Hills, CA: Sage Publications.

Kim, J.-O., & Mueller, C. W. (1986b). Introduction to Factor Analysis - What It Is and How to Do It. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-013). Beverly Hills, CA: Sage Publications.

Krasner, S. D. (1983). Structural Causes and Regime Consequences: Regimes as Intervening Variables. In S. D. Krasner (Ed.), International Regimes, (pp. 1-21). Ithaca: Cornell University Press.

Kruskal, J. B., & Wish, M. (1986). Multidimensional Scaling. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-011). Beverly Hills, CA: Sage Publications.

McIver, J. P., & Carmines, E. G. (1981). Unidimensional Scaling. (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-024). Beverly Hills, CA: Sage Publications.

Most, B. A., & Starr, H. (1989). Inquiry, Logic and International Politics. Columbia, S.C.: University of South Carolina Press.

Ness, G. D. (1985). Managing Not-So-Small Numbers - Between Comparative and Statistical Methods. International Journal of Comparative Sociology, 26, 1-13.

Przeworski, A., & Teune, H. (1970). The Logic of Comparative Social Inquiry. New York, N.Y.: Wiley-Interscience.

Singer, J. D. (1969). The Level-of-Analysis Problem in International Relations. In J. N. Rosenau (Ed.), International Politics and Foreign Policy, (revised ed., pp. 20-29). New York: Free Press.

Small, M., & Singer, J. D. (1982). Resort to Arms - International and Civil Wars, 1816-1980. Beverly Hills, CA: Sage Publications.

Sprinz, D. (1993). The Impact of International and Domestic Factors on the Regulation of Acid Rain in Europe: Preliminary Findings. Journal of Environment & Development, 2(1), 37-61.

Sprinz, D., & Helm, C. (1996). International Governmental Organizations and National Participation in Environmental Regimes: The Organizational Components of the Acidification Regime (EU Contract EV5V-CT94-0390) (mimeo.). Potsdam: Potsdam Institute for Climate Impact Research.

Sprinz, D., & Vaahtoranta, T. (1994). The Interest-Based Explanation of International Environmental Policy. International Organization, 48(1), 77-105.

Sprinz, D. F. (1992). Why Countries Support International Environmental Agreements: The Regulation of Acid Rain in Europe. Unpublished Ph.D. diss, Ann Arbor, MI: Department of Political Science, The University of Michigan.

Sprinz, D. F. (forthcoming-a). Domestic Politics and European Acid Rain Regulation. In A. Underdal (Ed.), The International Politics of Environmental Management. Dordrecht: Kluwer Academic Publishers.

Sprinz, D. F. (forthcoming-b). Regulating the International Environment: A Conceptual Model of Environmental Security and Instrument Choice. In N. P. Gleditsch (Ed.), Conflict and the Environment: Proceedings of a NATO Advanced Research Workshop Held at Bolkesjø, Norway, 12-16 June 1996. Dordrecht: Kluwer.

The Economist. (1997). The Asian Miracle - Is It Over? The Economist, 342(8006), 23-25.

UNDP. (1994). Human Development Report 1994. New York, N.Y.: Oxford University Press.

Widmer, T. (1991). Evaluation von Massnahmen zur Luftreinhaltepolitik in der Schweiz - Eine quasi-experimentelle Interventionsanalyse nach dem Ansatz Box/Tiao [Evaluation of Clean Air Policies in Switzerland: A Quasi-Experimental Intervention Analysis Using the Box/Tiao Approach]. Zürich: Verlag Rüegger.

Wonnacott, R. J., & Wonnacott, T. H. (1979). Econometrics. (second ed.). New York, N.Y.: John Wiley & Sons.

Wonnacott, T. H., & Wonnacott, R. J. (1984). Introductory Statistics for Business and Economics. (third ed.). New York, NY: Wiley.

World Bank/Environmentally Sustainable Development. (1995). Monitoring Environmental Progress - A Report on Work in Progress. Washington, D.C.: International Bank for Reconstruction and Development.

World Resources Institute. (1996). World Resources 1996-97 - A Guide to the Global Environment: http://www.wri.org/wri/wr-96-97/.

---

[i] Qualitative studies may also incorporate multiple explanatory variables. Given the low number of cases relative to the number of explanatory variables, inferences regarding the effect of each explanatory factors as well as their joint impact become difficult to assess in empirical-qualitative research.

[ii] Experiments use *random* assignment of cases to treatment and control groups; therefore, the confidence in findings from experimental research is generally higher than for any other type of research design.

[iii] An excellent and concise introduction to the general linear model can be found in Darnell (1994, 158-160).

[iv] An excellent treatment of necessary and sufficient conditions can be found in Most and Starr (1989, ch. 3).

[v] For a definition of construct validity, see the section on "Threats to Validity" further below.

[vi] For readers interested in the literature on scaling, the publications by Kruskal and Wish (1986), Jacoby (1991), McIver and Carmines (1981), as well as Kim and Mueller (1986a; 1986b) provide points of departure.

[vii] Introductions to a broad variety of methods of statistical analysis can be found in standard textbooks (e.g., Greene, 1997; Judge, Griffiths, Hill, Lütkepohl & Lee, 1985; Wonnacott & Wonnacott, 1979; Wonnacott & Wonnacott, 1984). A state-of-the-art review of the use of advanced statistical techniques in Political Science can be found in Bartels and Brady (1993).

[viii] Excellent summer courses in the field of political science methodology are conducted, inter alias, at the Inter-University Consortium for Political and Social Research at the University of Michigan, Ann Arbor, MI, USA (http://icpsr.umich.edu) and the European Consortium for Political and Social Research at Essex University, Essex, U.K. (http://www.essex.ac.uk/ECPR).

[ix] For a detailed treatment of threats to validity, see Cook and Campbell (1979, ch. 2).

[x] In illustrating the accomplishments and pitfalls in quantitative analyses of international environmental policy, I will disproportionally draw on the research by Jänicke and his group of collaborators. The selection bias is explained by the fact that this group of researchers has the longest track record of published research, and their explicit documentation permits a detailed critique.

However, this shall not lead readers to insufficiently cherish the contributions which this group of researches has made to the field. Most likely, research by others suffers from the same problems, but it is either not yet published or not as well documented.

[xi] For a general strategy to deal with medium-sized N datasets, see Ness (1985).

[xii] This method reduces the number of variables by postulating that one variable out of a group of conceptually related variables accounts for the hypothesized variation in the dependent variable.

[xiii] The country studies on the effect of structural changes on the economies uses (but does not systematically analyze) time series data (Jänicke et al., 1993), however, time series data appear not to be available for the crossnational analyses of environmental performance (ibid.).

[xiv] This is obviously rarely the fault of researchers as pollution inventories are undergoing revisions and data for longer time periods may not be available. The censoring of the time domain may, however, have substantive impacts on the fundamental results of research with the potential of providing inadequate guidance for public policy.

[xv] Heteroskedasticity is the absence of a constant discrepancy between observed and predicted values (i.e., non-constant error terms). In Ordinary Least Squares (OLS) estimation procedures, this leads, inter alias, to biased estimations of the variance of parameter coefficients. Generalized Least Squared (GLS) provide an appropriate alternative to OLS in such cases (Kennedy, 1992, 114-116).

[xvi] It may be worth noting that there are no "spurious correlations." While the substantive interpretation of a relationship by the researcher may be spurious, measures of associations are results of software algorithms (the latter may, at times, be spurious due to programming errors).

[xvii] For an illustrative example, see The Economist (1997).

[xviii] For details, please consult their WWW site at http://www.ciesin.org.