Nested Analysis as a Mixed-Method Strategy for Comparative Research

EVAN S. LIEBERMAN *Princeton University*

Despite repeated calls for the use of "mixed methods" in comparative analysis, political scientists have few systematic guides for carrying out such work. This paper details a unified approach which joins intensive case-study analysis with statistical analysis. Not only are the advantages of each approach combined, but also there is a synergistic value to the nested research design: for example, statistical analyses can guide case selection for in-depth research, provide direction for more focused case studies and comparisons, and be used to provide additional tests of hypotheses generated from small-N research. Small-N analyses can be used to assess the plausibility of observed statistical relationships between variables, to generate theoretical insights from outlier and other cases, and to develop better measurement strategies. This integrated strategy improves the prospects of making valid causal inferences in cross-national and other forms of comparative research by drawing on the distinct strengths of two important approaches.

Long-standing methodological debates highlighting inherent tradeoffs in the main modes of comparative analysis have tended to force scholars to choose between one of two imperfect approaches. On the one hand, even while defending its merits, Lijphart (1971, 685) succinctly identified the central shortcoming of the "comparative method" as the problem of "many variables, small number of cases." In the years to follow, some scholars argued that such attempts to draw general conclusions from intensive analysis of one or a few cases have been flawed by various problems of selection bias, lack of systematic procedures, and inattention to rival explanations (e.g., Achen and Snidal 1989; Geddes 1990; King, Keohane, and Verba 1994). Alternatively, other scholars have argued not only that some of the critiques of qualitative research may be overdrawn and the contributions of these works underappreciated, but also that the complex phenomena and causal processes associated with big, national-level outcomes require a more close-range analytic tool that is less likely to generate spurious results (e.g., Collier, Brady, and Seawright 2004; Collier and Mahoney 1996; Munck 1998; Rogowski 1995). Qualitatively oriented scholars have their own tradition of challenging the statistical approach, including Sartori's (1970) powerful invective against "conceptual stretching," which in turn has been refuted by scholars such as Jackman (1985), who argues that the comparative method is a "weak approximation of the statistical method," (165) and that "cross-national statistical analyses have a lot to offer" (179).

Although such back-and-forth debate has served to illuminate the shortcomings in various methodological approaches, it has also provided momentum for greater synthesis of research styles and findings. Two decades after publication of Lijphart's (1971) article, Collier (1991) pointed out that advances in both statistical and small-N approaches, and evidence of increasing communication across the two approaches, held great promise for scholarly progress. Both King, Keohane, and Verba's *Designing Social Inquiry* (1994) and Brady and Collier's *Rethinking Social Inquiry* (2004) have demonstrated that each mode of analysis can be successfully used to achieve similar social scientific ends, while using somewhat different tools. And yet, these contributions have largely assumed that there will continue to be substantial divisions of scholarly labor, even as research findings across the methodological divide are often ignored. In a somewhat different formulation, several scholars have called for greater *integration* of methodological approaches (Achen and Snidal 1989; Tarrow 1995) or the mixing of methods. Despite the initially appealing nature of such a resolution, scholars have received little guidance about how to blend these modes of analysis. As Bennett (2002) points out in a paper reviewing some of the ways in which case study, statistical, and formal methods have been combined in political science, there is a need to focus on the ways in which such combinations could be increased and improved. Clearly, not all forms of mixed strategies will provide greater insights into particular research problems. In fact, some may simply generate more confusion than clarity.

This article systematizes a unified "mixed method" approach to comparative research, which I call *nested analysis*.[1] It combines the statistical analysis of a large

Evan S. Lieberman is Assistant Professor of Politics at Princeton University, 239 Corwin Hall, Princeton, NJ 08544 (ESL@Princeton.edu).

Earlier versions of this paper were presented at the 2002 Annual Meeting of the American Political Science Association, Boston, MA, and at a 2004 Workshop on multilevel models at Princeton University. In addition, Larry Bartels, Andrew Bennett, Nancy Bermeo, David Collier, Michael Coppedge, John Gerring, Marc Morjé Howard, David Laitin, Phil Shively, students in my Qualitative Methods graduate seminars at Princeton, members of the Princeton Comparative Politics Working Group, three anonymous reviewers and the editor of the *APSR* all provided extremely helpful comments and suggestions.

[1] In this article, I discuss Coppedge's (2005) use of what he calls "nested inference" in an analysis of the breakdown of democracy in Venezuela. Although he is methodologically self-conscious in describing how case study and quantitative/large sample analyses are combined in an application, that study represents one variation of the approach I describe in this article, which attempts to anticipate and systematize a broader range of research problems and strategies.

sample of cases with the in-depth investigation of one or more of the cases contained within the large sample. This would include the study of a nation-state nested within an analysis of 50 nation-states; the study of two provinces nested within an analysis of 20 provinces; or the study of an institution nested within an analysis of 100 institutions. Although all of the examples discussed in the article are concerned with country- or national-level analyses, the strategies described here should apply to any comparative analysis of social units for which both quantitative and in-depth case study data can be obtained. Thus, the approach could be applied to the analysis of individual behaviors or attitudes, but only if the researcher were willing and able to gather new data about particular individuals through intensive interview or related approaches in combination with quantitative analyses of large-scale surveys. If the study concerned specific, well-studied individuals, such as presidents or legislators, for which additional information could be gleaned from in-depth research of particular cases, the approach described here would indeed apply.[2]

I should be clear that the strategy described here is quite distinct from the message outlined by King, Keohane, and Verba (1994). Rather than advocating that there are "lessons" useful for qualitative researchers that can be gleaned from the logic of statistical analysis (or *vice-versa*, an argument they do not make) I show that there are specific benefits to be gained by deploying *both* analytical tools simultaneously, and I emphasize the benefits of distinct complementarities rather than advocating a single style of research. Although the move from "small-N" analysis (SNA) to nested analysis obviously requires that one "find additional cases" (King, Keohane, and Verba 1994, 208–29), it assumes that it may be extremely difficult and inefficient to gather perfectly equivalent data for each case, and that the inferential opportunities from the "large-N" analysis (LNA) will be distinctive.[3]

OVERVIEW OF THE NESTED ANALYSIS APPROACH

I describe a set of strategies for gaining maximum analytic leverage when combining SNA and LNA within a single framework (summarized in Figure 1). Although there is an enormous variety of analytical strategies contained under these two subheadings, both in terms of actual number of units analyzed and the scope of the time dimension considered, for the purposes of this paper, it is useful to make a general distinction: I define LNA as a mode of analysis in which the primary causal inferences are derived from statistical analyses which ultimately lead to quantitative estimates of the robustness of a theoretical model; I define SNA as a mode of analysis in which causal inferences about the primary unit under investigation are derived from qualitative comparisons of cases and/or process tracing of causal chains within cases across time, and in which the relationship between theory and facts is captured largely in narrative form.[4] The strategy of combining the two approaches aims to improve the quality of conceptualization and measurement, analysis of rival explanations, and overall confidence in the central findings of a study. The promise of the nested research design is that both LNA and SNA can inform each other to the extent that the analytic payoff is greater than the sum of the parts. Not only is the information gleaned complementary, but also each step of the analysis provides direction for approaching the next step. Most prominently, LNA provides insights about rival explanations and helps to motivate case selection strategies for SNA, whereas SNA helps to improve the quality of measurement instruments and model specifications used in the LNA.

As a thumbnail sketch, the approach involves starting with a preliminary LNA and making an assessment of the robustness of those results. If the model is well specified and the results are robust, one proceeds to "Model-testing Small-N Analysis," and if not, to "Model-building Small-N Analysis." In each case, as shown in Figure 1, the analyst must again make assessments about the findings from such analysis, using directions and insights gleaned from the SNA, and those assessments provide a framework for either ending the analysis or carrying out additional iterations of SNA or LNA. Detailing the nature of the particular strategies for carrying out each type of analysis, as well as the nature of the assessments, is the central goal of the remainder of the paper.

Nested analysis is resolutely "catholic" in its assumptions and objectives. It assumes an interest in *both* the exploration of general relationships and explanations and the specific explanations of individual cases and groups of cases. For example, a nested research design implies that scholars will pose questions in forms such as "What causes social revolutions?," while simultaneously asking questions such as "What was the cause of social revolution in France?" Nested analysis helps scholars to ask good questions when analyzing their data and to be resourceful in finding answers.

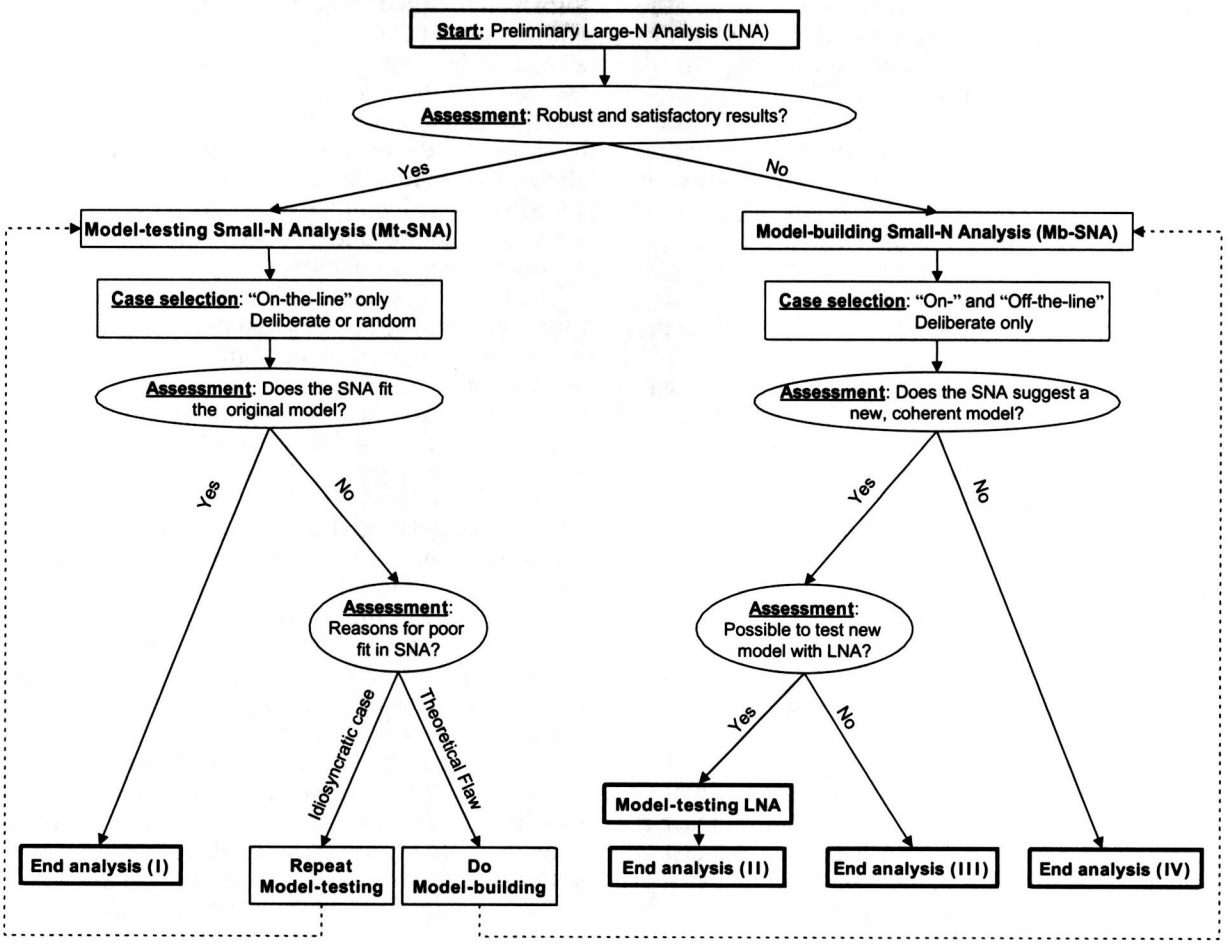
Before proceeding to detail the procedures associated with nested analysis, it is important to situate the strategy within the context of two other proposals for "alternative" methodological approaches. First, Charles Ragin (1987, 2000) attempts to steer a middle path between "quantitative" and "qualitative"

[2] However, for most analyses of individual behaviors or attitudes, for which the "large-N" component of the data is contained in a survey, I would not expect this approach to be feasible, because scholars are unlikely to be able to conduct further in-depth research with the original respondents. Moreover, the prospect of explaining the exceptional nature of a particular individual is unlikely to be of intrinsic interest in the way scholars are likely to be interested in the particularities of larger social units, such as national states.

[3] By "cases" I mean the shared unit of analysis. In cross-national research, each case is a country.

[4] As is discussed in the text, qualitative analysis is the hallmark, but not the defining feature of SNA. Within-case analyses may include a range of statistical analyses of data that are not available across the larger sample of primary unit cases (i.e., countries).

FIGURE 1. Overview of the Nested Analysis Approach



research with his specification of a Boolean approach and elaboration of a "fuzzy set"/Qualitative Comparative Analysis (FsQCA). Ultimately, his strategy focuses on integrating close-range analysis to ensure the proper delineation of theoretically relevant populations and valid classification of cases, with an algorithm that finds the necessary and sufficient conditions associated with particular sets of phenomena. Second, the Bayesian approach (Western and Jackman 1994), like the FsQCA approach, and distinct from the classical regression model, relies heavily on investigator knowledge of cases and processes, but does this through the formal introduction of subjective probability estimates. However, neither the stated approaches to Bayesian analysis nor FsQCA provide direction about how to gather additional research in the SNA—they assume a seamless discovery process of "outside knowledge," with almost no focus on the specific role of gathering and reporting case materials. In making prescient critiques of standard cross-country regression analyses, advocates of both the Bayesian and the FsQCA approaches allow for the inductive incorporation of knowledge from cases, but as currently formulated, they provide little guidance about the cases we should study or what role they ought to play in the assessment of theoretical findings.[5] As such, both of these approaches may serve as partial correctives to cross-country regression analysis, but neither is complete. For the purposes of nested analysis, both FsQCA and Bayesian approaches may be used in the LNA, and the guidelines developed here for combining such approaches with SNA should still apply.

It is also important to indicate that the nested analysis approach is agnostic with respect to the source of theory formation. Although others have explicitly included the development of formal—that is, mathematically specified—theory in their discussions and proposals for integrating approaches to the study of comparative politics (Bates et al. 1998; Laitin 2002), the nested analysis approach has no particular affinity for any single theoretical approach, except for a more general positivist goal of causal inference. Such theory may be developed and conveyed in a nonmathematical form (i.e., "No bourgeoisie, no democracy") or through the use of mathematical operators and proofs. Along these lines, the nested analysis approach allows for both

[5] Certainly, the nested analysis approach could be described as a "folk Bayesian" approach (McKeown 2004, 158–62) in that it seeks to formally introduce investigator knowledge of the world.

the testing of deductively formed hypotheses and the inductive generation of theory. Many of the benefits of nested analysis explicitly rest on the assessment that the overall state of theory in cross-national research is relatively thin with respect to the questions being asked, and that empirical analysis is required both to develop hypotheses and to test them. Of course, the nature of the specific hypotheses—including the reliance on microfoundational or macrostructural mechanisms—is likely to shape the evidentiary requirements, particularly in the SNA (as discussed later). In the remainder of the article, my use of the term *model* implies only a general theoretical argument that relates explanatory to outcome variables, and should not be taken to imply a "formal" model.

The central objective of the remainder of the article is to specify a set of procedures for integrating LNAs and SNAs. Although it is neither possible nor desirable to identify a cookie-cutter approach to analysis, the systematization of these steps should provide a clear logic for integrating the two types of analyses and for identifying the types of assumptions and justifications that are required for analysis. As always, scholarly tastes and subjective judgments about the robustness of the results influence how the nested analysis will proceed,[6] but it is important to ensure a high degree of transparency, particularly when adding complexity to the scope of analysis.

I use examples of published and unpublished studies to demonstrate the use of various techniques within the larger approach of nested analysis, but the article is not intended to be a review of the literature as much as an outline for the execution of nested analysis. Indeed, because almost none of the examples actually employ the specific language or framework developed here, I only claim that these examples help to clarify how aspects of the approach have been used in particular studies and with what benefit.

STARTING THE ANALYSIS: PRELIMINARY LNA

Scholars engage new research projects with varying levels of background information about a specific case or set of cases, but the nested analysis formally *begins* with a quantitative analysis, or preliminary LNA. Thus, a prerequisite for carrying out a nested analysis is availability of a quantitative dataset, with a sufficient number of observations for statistical analysis,[7] and a baseline theory. The preliminary LNA provides information that should ultimately complement the findings of the SNA, and that will guide the execution of the SNA. Particularly for scholars who would have carried out SNA exclusively, the preliminary LNA requires explicit consideration of the universe of cases for which the theory ought to apply, and identification of the range of variation on the dependent variable. It also provides opportunities to generate clear baseline estimates of the strength of the relationship between variables, including estimates of how confident we can be about those relationships given a set of assumptions about probabilities and frequencies. When scholars begin with strong hypotheses and good data, the preliminary LNA can be understood as a more conventional hypothesis-testing analysis.

The content of the LNA may take one of several forms, depending on the availability of data, and the nature of the causal model—for example, depending on whether the outcome is understood to be graded or dichotomous, and whether the hypothesized relationship is understood in probabilistic or deterministic terms. One may use multivariate regression analysis; fuzzy set/qualitative comparative analysis (FsQCA); bivariate/correlational analysis, or simply descriptive statistics to analyze the scores on the dependent variable. Decisions about which brand of analysis to use, and the nature of the model—linear or curvilinear, for example—must be made with respect to available data and theory. In any case, the goal of the preliminary LNA is to explore as many appropriate, testable hypotheses as is possible with available theory and data. Indeed, the very feasibility of nested analysis is a product of the increasing availability of datasets produced by other scholars and international organizations, obviating the need for independent data collection, at least at this preliminary stage. (Significant independent collection of data at this stage can be justified only when a scholar has very strong initial hypotheses and great confidence in how to measure key variables.) However, it is important to note that the preliminary LNA should avoid the insertion of any control variables that do not have a clear theoretical justification, such as regional "dummy" variables. Such variables *are* likely to soak up some of the cross-country variance, leaving less to be explained in the SNA, but in the absence of good theory, such controls weigh against the nested approach, which aims to answer the very question of why groups of countries might vary in systematic ways.

A core strength of LNA relative to SNA is its ability to simultaneously estimate the effects of rival explanations and/or control variables on an outcome of interest. To a large extent, SNA in the field of comparative politics has relied on variants of Mill's methods in order to deal with country-level rival explanations—that is, scholars identify cases that score similarly on several key variables, using shared traits as a basis for analytical equivalence approximating statistical control.[8] Although the strategy of identifying cases with relatively similar scores on such variables can be a powerful one, in a nonexperimental setting, important differences among cases can almost always be identified,

[6] Indeed, there is no consensus about the robustness of a particular R^2 statistic, or what amount of process-tracing evidence should be considered persuasive.

[7] There is no clear lower bound for the number of cases that can be analyzed through a statistical analysis, but fewer cases obviously reduce the degrees of freedom and intrinsic power of the analysis. It is rare to see quantitative analyses of fewer than 12 cases in cross-country regression analyses.

[8] See Gerring 2001 (209–14) for a summary of these methods, often understood as "most similar" and "most different" systems research designs.

and these emerge as possible rival explanations. Regardless of whether one's causal model is probabilistic or deterministic, some degree of covariance between a rival explanatory variable and the outcome requires attention within a SNA based on the juxtaposition of "similar" cases. One may attempt to draw on theory to argue why a particular variable is an implausible influence, but skeptics are likely to demand empirical proof. Moreover, one may attempt to carry out "within-case" analysis (Collier 1999) to address rival hypotheses, but again, there may be no over-time variation or other relevant data to analyze; or, one may try to find an additional "similar" case with less variation on the offending variable, but in a world with a limited number of highly heterogeneous countries, such options may be limited. Alternatively, SNA scholars can ignore the cross-case variance or simply concede that there is no way to address the problem with available data. Obviously, these are not ideal solutions.

Depending on the question or the cases under investigation, LNA may be able to lend a hand. Assuming that the LNA is conducted as a regression, the relevant dependent variable can be regressed on measures of the rival explanatory variable under investigation in order to assess the strength of a relationship, particularly when the SNA provides no solid basis for analysis. For example, in her study of multilateral sanctions, Lisa Martin (1992) precedes her analysis of four major case studies with a set of regression analyses, which allows her to assess the general plausibility and implausibility of several possible explanations of why states cooperate to impose economic sanctions. She argues that this technique "has allowed us to narrow the range of hypotheses deserving more-detailed analysis by suggesting that some hypotheses... have little empirical support" (92). For example, she finds no support for Keohane's (1984) "declining hegemony" thesis in the LNA (91), which allows her to focus her attention on other possible explanations in the subsequent SNA. In the absence of such LNA, Martin would have been forced to consider Keohane's (1984) important hypothesis in the SNA, imposing analytic costs, and leaving readers to wonder about the weight of this explanation in the larger sample. Alternatively, if the LNA had provided initial support for the Keohane thesis, Martin either would have been forced to accept the usefulness of that model—and perhaps demonstrate that other complementary explanations were possible—or would have been forced to demonstrate in quite convincing terms within the SNA why statistical relationships were likely spurious. Clearly the most powerful refutation of a rival explanation is the presentation of disconfirming evidence in *both* LNA and SNA, but given data and analytic constraints, the ability to rule out a hypothesis in the LNA provides sound justification for focusing on other explanations in the SNA.

At least as important as its ability to dismiss rival explanations, LNA provides a unique instrument for assessing the strength of partial explanations or control variables. Because country-level outcomes tend to be the product of several factors, preliminary LNA is likely to find that some variables *are* significant predictors of the outcome under investigation, even if they can account for only a limited portion of the cross-country variance. For example, in a study of the development of the tax state, Lieberman (2003) begins his analysis by demonstrating that approximately 40% of the cross-national variation in levels of income tax collections is predicted by levels of GDP/capita. Although this variable is essentially treated as a control variable throughout the book, it is extremely useful to have an estimate of the extent to which such a variable helps to explain patterns of variation on the outcome. Much small-N research involves the comparison of "similar" cases. However, because we only observe cases in which there is little to no variation on key control variables, we have little basis for making inferences about the need to control on those variables, or about how strong an influence we should expect those variables to have on the outcomes under investigation. The "puzzle" of a particular case or set of cases can be made clear when we have some estimate of predicted outcomes given a set of parameter estimates and the case scores on those variables.[9]

ASSESSING THE FINDINGS OF THE LNA: ARE THE RESULTS ROBUST AND SATISFACTORY?

Beyond providing insights into the range of variation on the dependent variable, and estimates of the strength of rival hypotheses and control variables, LNA also provides important information about how to carry out the next stage of the analysis—intensive examination of one or more cases. First, the scholar must assess the findings: did the preliminary LNA provide strong grounds for believing that the initial theoretical model explained the phenomenon being studied?

As noted previously, it is not possible to provide absolute criteria for answering the question about the robustness of the LNA results because subjective assessments about the state of knowledge and what constitutes strong evidence weigh heavily.[10] Depending on the nature of the LNA, standard assessments about the strength of parameter estimates must be used to evaluate goodness of fit between the specified model and the empirical data. Nonetheless, one important tool is central to the nested analysis approach: the actual scores of the cases should be plotted graphically relative to the predicted scores from the statistical estimate,[11] and with proper names attached. This provides an opportunity to make specific assessments of the goodness

[9] Although it is true that these initial parameter estimates are likely to be biased because of model misspecifications, including missing variable bias, our presumption is that when we do not have a fully specified or complete theoretical model, it is useful to gain a sense of what can be explained by the theory and data that *are* available.

[10] For a classic statement on the use of common sense and professional judgment in the use of quantitative analysis, see Achen (1982), especially pp. 29–30.

[11] At the extreme, if no statistical relationship is found between any of the explanatory variables and the outcome of interest, one could simply use a central tendency of the data, such as the mean, as a baseline model, and country cases could be plotted as deviations from the mean.

of model fit with the available cases. In combination with the parameter estimates generated from the LNA, the scholar must decide if the unexplained variance is largely the product of random noise, or if there is reason to believe that a better model/explanation could be formulated. As in any statistical analysis, diagnostic plots may highlight suspect patterns of nonrandom variation in one or more cases—the identification of outliers. However, unlike in surveys of individuals, where case identities are anonymous and thus irrelevant for analysis, in the study of nation-states and many other organization forms, the location of specific cases with respect to the regression line may strongly influence one's satisfaction with the model. For example, a scholar may feel unsatisfied with a model that cannot explain a case perceived to be of great significance within the scholarly literature (e.g., the French revolution in the study of revolutions), or the identification of an outlier case may immediately suggest a new theoretical specification with potentially broader application. If a scholar enters the research project with specific hunches about seemingly anomalous outcomes, analysis of the actual-versus-predicted-scores plot may demonstrate that one or more cases are indeed outliers that may warrant more theoretical attention. Indeed, Lieberman's (2003) study was motivated by a hunch that differences in the Brazilian and South African tax structures were striking and not readily explainable, and the preliminary LNA confirmed that this was true even when key control variables were taken into account. Of course, such preliminary analysis could have served to foreclose unnecessary research by demonstrating that a particular case was (surprisingly) well explained by the existing state of theory.

Using such analyses, the scholar must answer the question: "Were all of the most important hypotheses tested and were the results robust/satisfactory?" The answer to this question informs the approach to the nested case analyses, or SNA, as described in the following section.

NESTING INTENSIVE CASE STUDIES (SNA) INTO THE ANALYSIS

The second major step of the nested analysis involves the intensive analysis of one or more country cases.[12] Of course, there is nothing particularly distinctive about the simple combination of LNA and SNA; scholars have long recognized the value of "triangulation" for descriptive and causal inference.[13] My contention is that there are several important strategies that can be gleaned from assessment of the LNA, which will narrow the larger menu of options for executing the SNA.[14] Moreover, I emphasize that the best use of SNA is to leverage its distinct complementarities with LNA, not to try to implement it with the exact same procedures as one would carry out regression analysis. Although many small-N scholars may have an "implicit" regression model in their head when they carry out their analyses, there are clear benefits to being explicit.[15]

It is important to recall that the goal of a nested analysis is ultimately to make inferences about the unit of analysis that is shared between the two types of analysis—typically countries or country-periods. In pursuing this goal, a nested analysis requires a *shifting* of levels of analysis because the SNA component demands an examination of *within*-case processes and/or variation.[16] The SNA should be used to answer those questions left open by the LNA—either because there were insufficient data to assess statistical relationships or because the nature of causal order could not be confidently inferred. For example, in a hypothetical study of the determinants of government policy, in which the LNA confirmed a hypothesized relationship between institutional form and policy outcome, the SNA would likely investigate the specific actions of groups and/or individuals within a given country. This would be done in an attempt to find specific evidence that the patterns of human organization hypothesized to have been influenced by the institutional form were actually manifest in reality. Moreover, the SNA is particularly useful for investigating the impact of rival explanations for which we lack good cross-country data.

The synergistic qualities of LNA and SNA reflect the different types of data that each brings to the analysis of a problem, and their relative strengths in the task of causal inference. Here it is extremely useful to highlight the distinction between a "data-set observation," which corresponds to a row in a rectangular dataset, and a "causal-process observation," which is "the foundation for process-oriented causal inference. (It) provides information about mechanism and context" (Collier, Brady, and Seawright 2004, 253). We can say that LNA is, by definition, comprised only of dataset observations, whereas the hallmark of SNA is a much smaller number of dataset observations *and* a host of causal process observations.[17] Within-case analysis generally entails the scrutiny of a heterogeneous set of materials,

[12] SNA involves multiple "within-case" observations, across space, time, and/or other dimensions. LNA may also involve multiple observations of country cases when cross-sectional data are pooled across time.

[13] See, for example, Ragin (1987, 69–84) for an excellent analysis of several combined approaches.

[14] Just as there are many styles and strategies for statistical analysis, there are at least as many approaches to SNA—an approach that, almost by necessity, involves less methodological structure than LNA because the analysis is strongly oriented toward discovery of novel social and political processes that take place in distinctly different ways across time and space. In recent years, there has been increasing methodological attention to the different types of strategies used by scholars when studying one or a few cases intensively. However, echoing the statements made previously with respect to LNA, this is not the place to review all of the distinctions about how such work is carried out. See, for example, contributions in Mahoney and Rueschemeyer 2003, Brady and Collier 2004, and George and Bennett (2005).

[15] Thanks to Phil Shively for highlighting this central point.

[16] See Gerring (2004) for a discussion of within-unit analysis in case studies.

[17] The number of rows in a dataset is typically understood as the number of country cases, or "N," that distinguishes small-N and large-N research. By now, most methodologists agree that a small-N study will have many observations, but as Collier, Brady, and Seawright (2004) point out, different inferential strategies are used to

including printed documents, interviews, and other observations that provide important information about the social phenomena we seek to understand. Because such materials are produced in such different shapes and forms across time and space, it is often impossible to specify, *a priori*, a set of very precise coding rules that would allow for an easily repeatable data collection and analysis process. These materials provide more fine-grained measurements of a host of events and behaviors, at both the micro- and macrolevels, and often in close temporal proximity to one another. Such data are virtually impossible to capture across large numbers of countries in a consistent manner. Scholars gain analytic leverage when they scrutinize the theoretical implications of these observations, either by testing existing hypotheses or by inductively developing new propositions about general relationships between causes and effects.

Although the distinction between LNA and SNA is generally between quantitative and qualitative modes of analysis, some aspects of SNA may involve quantitative analyses at *different levels of analysis*. For example, one could analyze a survey of individuals for a given country if that analysis could shed light on the dynamics of the social or political process being studied for the country at large. Analyses of individual behavior are specifically relevant to the nested approach only to the extent that they shed light on the larger questions being considered in the LNA. In a similar manner, the SNA might include time-series analysis (using country-year as the unit of observation) as a way of linking cause to effect or for dealing with case-specific rival explanations, particularly when the LNA was carried out as cross-sectional analysis. For example, in Lieberman (2003), time-series analyses of the production of government tax collections in the SNA of South Africa helped to rule out the rival explanatory power of the role of early reliance on mining revenues, which would not have been possible in the cross-country LNA, for which comparable data were not available.

The inclusion of additional theoretically valid cases is always preferred in LNA, but practical constraints on investigator skills and time as well as the desirability and feasibility of reporting in-depth analyses on multiple cases create important tradeoffs which must be weighed by scholars when selecting cases for the SNA. There is no theoretical benchmark akin to probability theory that small-N scholars can draw on to establish precise guidelines about what constitutes compelling evidence. The very nature of "causal process observations" is that they are highly heterogeneous: some documented observations may serve as particularly powerful "smoking gun" evidence linking cause to effect, whereas others may simply serve as incremental steps that increase the plausibility of a set of theoretical claims. Small-N analysis provides the opportunity to implement various "quasi-experimental" explorations by looking at the impact of various shocks or treatments within the historical record.[18]

Particularly if one were to follow the recommendations of King, Keohane, and Verba (1994) to increase the number of observations, scholars might incorrectly conclude that the best strategy for the SNA component of the nested analysis would be to analyze as many country cases as possible. On the contrary, such a strategy tends to lead to a diminution of the core strengths of the SNA. Increased degrees of freedom are provided by the LNA, and nested analysis should rely on the SNA component to provide more depth than breadth—that is, given a fixed amount of scholarly resources, more energy ought to be devoted to identifying and analyzing causal process observations within cases, rather than to providing thinner insights about more cases. Because the inherent weakness of SNA is its inability to assess external validity, there is no point in trying to force it do this when the LNA component of the research design can do that work. Notwithstanding this advice, it will almost always be useful to evaluate more than one case in the SNA; the elaboration of concepts and mechanisms can best be accomplished through *comparison*. A great strength of small-N analysis is the juxtaposition of both similar and contrasting cases, helping to make transparent the operationalization of concepts that are largely hidden in the analysis of a statistical dataset. Furthermore, comparison provides an empirical basis for making narrative assessments of counter-factual claims—that is, an event would have happened a different way had the score on a key variable or set of variables been different (George 1979).

To the extent that scholars increasingly employ variants of nested analysis, standards will need to be established as to what constitutes an actual "case" study. For example, in studies that report statistical and case study findings, Reiter and Stam (2002), and Huth (1996) deploy what can be described as "mini-case analyses." These help to alert readers to examples of the argument being made by highlighting how well-known cases fit within their typologies and the degree to which they confirm to theoretical expectations. However, in these examples, the use of SNA is rather limited, and so little additional analytic value is gained. In these studies, the case analyses provide proper names for the independent and dependent variable scores, but they do not provide much elaboration about the alternative ways in which these scores were measured in comparison to the measurement procedures followed in the large-N dataset. Moreover, Reiter and Stam and Huth do not proceed with process tracing, linking cause to effect with any significant narrative. Just as statistical analyses

interpret such data. It is worth noting that even with these additional observations, such research is dubbed small-N—a convention that I use here. Meanwhile, the proliferation of TSCS analyses of country-level data is widely touted as useful strategies for increasing analytic power through a larger N (e.g., Beck and Katz 1995), but as Western and Jackman (1994, 414–5) observe, the time-invariant quality of many variables considered in cross-country analyses often implies that TSCS adds minimal additional analytic leverage for the overall problem being studied.

[18] See Campbell and Stanley 1966. I develop a set of strategies for exploring institutional hypotheses in small-N cross-country research in Lieberman 2001a.

must report on the sample size of the dataset, SNA demands full and clear exposition of the array of sources consulted and the depth of the historical analysis considered prior to writing the narrative.[19] As the number of cases in the SNA increases, the individual case analyses are likely to become increasingly superficial, and the distinct advantages of SNA are likely to diminish.

Beyond emphasizing the general complementarities, it is also important to focus the SNA based on the specific findings and analysis of the LNA. Recalling the question posed at the end of the previous section—namely, the analyst's assessment of the robustness of the preliminary LNA—SNA will then proceed along one of two tracks. If the answer is "yes, the results were robust," as indicated in Figure 1, then the goal of the SNA will be almost exclusively focused on *testing* the model estimated in the LNA. On the other hand, if the findings were not deemed to be robust, or if one or more important hypotheses could not be explored, including if the analyst believes that the appropriate theoretical model has not yet been specified, the SNA will be oriented toward model *building*. As I detail in the sections that follow, the decision about whether to proceed with a model-testing Small-N Analysis (Mt-SNA) or a model-building Small-N Analysis (Mb-SNA) will inform the scope of the analysis, the case selection strategy, and the analysis-ending criteria for the SNA. Practitioners may respond that SNA is itself a mix of model building and model testing and that the dichotomy is a false one. Although it is true that these may be "ideal-type" approaches, there is enormous benefit to being self-conscious about the *central* intention of one's research in the SNA stage, particularly because the nested approach provides distinct sets of guidelines for the respective strategies. Assessment of the preliminary LNA constitutes an important decision-point in how the nested approach will be carried out (as depicted in Figure 1), providing important guidelines for an appropriate analytic scope for the SNA.

Model-Testing SNA (Mt-SNA)

When scholars decide they are content with both the specification and fit of the model specified in the LNA, the main goal of the in-depth component of the nested research design is to further test the robustness of those findings. Given the potential for problems of endogeneity and poor data in statistical analyses carried out at the country level of analysis, statistical results alone rarely provide sufficient evidence of the robustness of a theoretical model. Almost inevitably, strong questions arise about causal order, heterogeneity of cases, and the quality of measurement. SNA provides an important opportunity to counter such charges. As Achen and Snidal (1989, 168–69) point out in an article otherwise quite critical of how such work is often practiced, "Case studies are an important complement to both theory-building and statistical investigations . . . they allow a close examination of historical sequences in the search for causal processes . . . Comparison of historical cases to theoretical predictions provides a sense of whether the theoretical story is compelling."

As the goal is to complement the LNA, the use of SNA in nested analysis should aim to gain contextually based evidence that a particular causal model or theory actually "worked" in the manner specified by the model. Can the start, end, and *intermediate* steps of the model be used to explain the behavior of real-world actors? Although this recommendation runs counter to the admonitions of Przeworski and Teune (1970), who argue that the ultimate goal should be to eliminate such labels, I believe that the nested analysis approach resonates more broadly with the general goals and expectations of scholars engaged in comparative research. That is, not only are we interested in our ability to make sense of patterns of variation, but also we would also like to use theory to account for decidedly important and seemingly anomalous outcomes in specific times and places. Moreover, unlike in some forms of medical research, where researchers are more likely to be content to find that a cause (say a drug used for minor pain relief) is related to a particular effect (say, better coronary health), even if they cannot identify the causal pathway of this relationship, social scientists are much less likely to be content with analogous findings. A good social science theory should not merely predict a particular relationship between independent and dependent variables, but it should also explain how and why these factors are related to one another (Gerring 2005), suggesting implications for what types of events and/or processes lie between cause and effect. SNA aims to make specific observations between those two points, verifying the plausibility of the stated mechanisms in terms of actions, outcomes, and/or perceptions. The SNA ought to demonstrate within the logic of a compelling narrative that in the *absence* of a particular cause, it would have been difficult to imagine the observed outcome.

In the case of Mt-SNA, scholars can justifiably focus their investigative resources on researching and analyzing the statistically significant results. The combination of theory and statistical results compels us to gather evidence—in the form of primary and secondary printed sources, interviews, surveys, and the other types of materials typically consulted for the development of an in-depth case analysis—that allows us to write a detailed narrative from the vantage-point of the preferred model. The evidence required for the SNA depends upon the nature of the theory. For instance, in a highly structural argument, actors may not be very aware of the circumstances that shape their actions, and so evidence of large-scale processes and events will be more appropriate than in the case of agent-oriented models, in which we would expect evidence of individual-level calculations and deliberate action.

While retaining a focus on assessing the plausibility of the preferred model, Mt-SNA should also aim to

[19] We should not establish as a standard for SNA that a longer narrative necessarily implies more careful research and/or analysis. Our assessment of the findings should be based on the methods used to gather and to analyze such data.

address two types of rival explanations.[20] First, if there were strong hypotheses that could not be considered in the LNA because of lack of cross-country data, the analyst should try to assess the strength of the hypothesis in the case study or studies. If cause and effect do not co-vary in the predicted manner, and/or if it is not possible to develop a coherent causal narrative guided by the rival model, the rival hypothesis can be dismissed. Second, the scholar should verify that the cause preceded the effect. Cross-country statistical databases (used in the LNA) are often highly limited in terms of temporal scope, and SNA can be used to verify that prior historical factors did not produce the observed result.

Model-Building SNA (Mb-SNA)

When the state of theory is initially weak or refuted by the LNA and/or the quality of the cross-country statistical data is not sufficient to adequately assess the chief hypotheses, the SNA will be called on to do more work. In this instance, the nested analysis approach demands a more wide-ranging and inductive Mb-SNA. Although scholars may initiate a research project with only general theoretical hunches, Mb-SNA involves using various case materials to develop well-specified theoretical accounts of cross-country variation on the outcome of interest. Moreover, the Mb-SNA ought to be used to identify measures that are valid and reliable indicators of the analytic constructs within the theoretical model. A clear shortcoming of LNA as it is often practiced in cross-country research is that many "off-the-shelf" datasets tend to be used, and variables may not actually measure what the theory describes.[21] Particularly in the instance of Mb-SNA, the investigator's proximity to a wide range of data sources should facilitate the development of valid measures.

As stated at the outset, many scholars may eschew the goal of identifying broadly generalizable theories or covering laws[22] and may use the LNA portion of the nested analysis approach simply to point out the limits of existing data and theory, motivating a more inductive search for explanations within a single case or small set of cases. Others seek more nomothetic findings. In either case, the scholar engaged in Mb-SNA does *not* proceed with the notion that a fully specified model is available and must develop explanations for the puzzle of varied outcomes. Although the Mt-SNA approach assumes that the refuted alternative hypotheses were adequately tested, the Mb-SNA approach invites reexamination of all theoretically strong propositions to the extent that data are available.

Inevitably, close-range analysis of one or a few country cases entails making difficult choices about which materials to investigate and which leads to pursue. Nevertheless, in most instances Mb-SNA has several advantages compared to SNA carried out in the absence of a preliminary LNA (i.e., a nonnested design). First, the scholar is equipped with useful, if partial, information about the strength of rival explanations and control variables. Of course, the reason for the negative results may be due to the poor quality of the data in the first place, but at the very least there is some indication about the weakness of relationships. Second, to the extent that the preliminary LNA provides a reasonable measure of the dependent variable, the Mb-SNA can focus on accounting for estimated differences between cases, or between cases and some central tendency of the population, having controlled for the effects of other influences. Third, the nested approach may induce the analyst to specify clearer concepts and models than conventional SNA, because even the anticipation of analyzing the results with statistical/quantitative tools implies the need for careful delineation of cause and effect. In this case, the SNA will be carried out with an eye toward theoretical parsimony and clarity, which are not always hallmarks of the SNA approach.

CASE SELECTION STRATEGIES FOR SNA

Nested analysis provides a solution to many of the tensions that exist in the current state of methodological advice about case selection strategies: scholars often justify intensive case study work because of a sense that they lack sufficient data and analysis of such cases, and yet most case selection strategies require that we justify that selection at the outset based on what we think we know about a particular case or set of cases, often in relation to a broader universe of cases. Nested analysis provides some assistance for squaring this circle, by detailing some guidelines for the daunting task of case selection with respect to the findings of the preliminary LNA. These strategies are useful when a scholar enters a research project without a prior inclination to investigate a particular case(s) and/or for assessing the analytic utility of certain case selection choices when a scholar already has a predisposition toward those cases prior to carrying out the preliminary LNA. Indeed, the nested analysis approach can leverage the accumulation of case-relevant skills and background (including language skills, case familiarity, etc.) which are important assets for most qualitatively oriented scholars. There is rarely a perfect case selection strategy for SNA. Rather, there is a set of options and choices that, again, may be narrowed significantly by the LNA. Specifically, we can make informed choices about whether to select cases based on predicted and actual scores on the independent or dependent variable and whether or not to select cases randomly.

[20] For a fuller discussion of the use of qualitative research to address rival explanations, see Collier, Brady, and Seawright 2004.

[21] See, for example, Lieberman 2001b for a discussion of how cross-country taxation data may (or may not) correspond with theoretical constructs about the relationships between state and society. Ragin (2000) makes an important point that the scale of country-level indicators (e.g., GDP/capita) may not correspond to differences in the underlying construct (e.g., level of development), and conceptually sensitive cutpoints and calibrations may be required.

[22] For a thoughtful challenge to the notion that comparative analysis should always involve the pursuit of nomological covering laws, see Zuckerman 1997.

Selecting Cases Relative to the Preliminary Model ("On-" or "Off-the-Line")

Perhaps no aspect of the methodological literature on case selection has left scholars engaged in small-N research more confused than the question of whether to select cases based on values on the independent or dependent variables. Particularly in the area of cross-national research, scholars have highlighted the pitfalls of selecting a case based on an extreme score on the dependent variable and attempting to infer general conclusions about the larger universe of cases (Geddes 1990). More recently, methodologists have highlighted a wider range of case selection options that will mitigate such problems, including the explicit accounting of the selection mechanism (King, Keohane, and Verba 1994, 128–37). More stridently, they recommend that scholars should select cases based on scores on the explanatory variable—a strategy that does not lead to analogous pitfalls of selection bias (King, Keohane, and Verba, 1994: 137–42)—while insisting that such cases be selected without knowledge of the dependent variable scores (142–46). Unfortunately, this solution, which attempts to replicate the inferential logic of experimental research, is largely impractical. In the first place, it assumes very strong theory, which is often not the case in cross-national research. In the second, because qualitatively oriented scholars tend to approach research questions from the perspective of trying to understand the determinants of puzzling outcomes, they are almost certain to know the scores on the outcome variable.

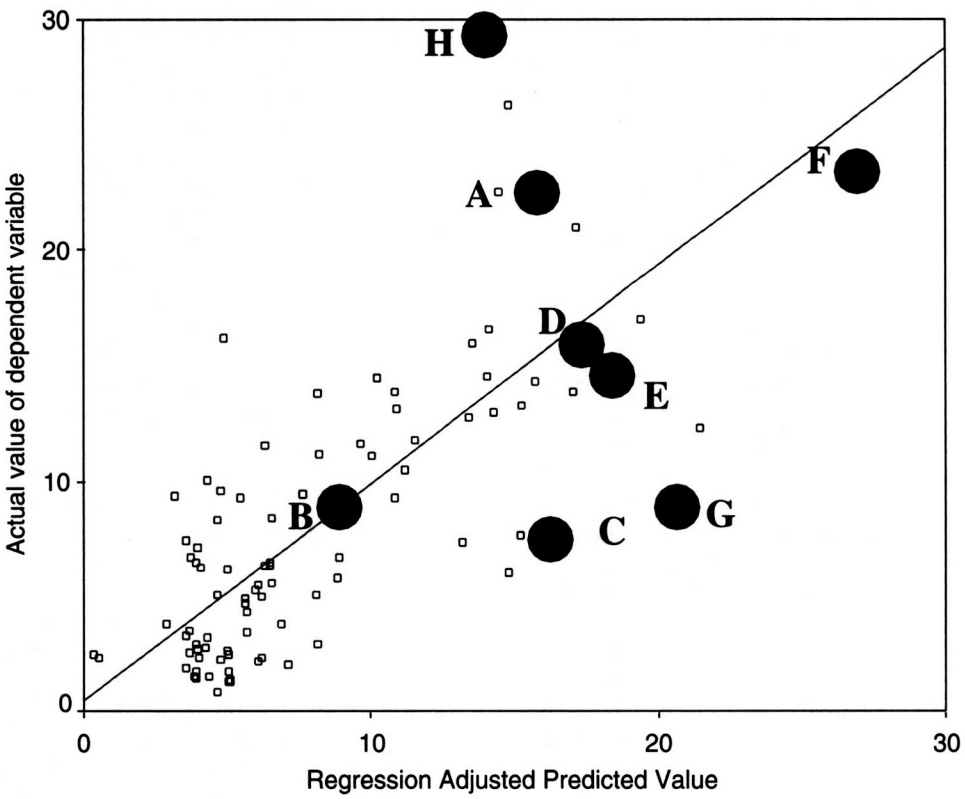
A second issue that comes up is whether we should investigate cases that are seemingly anomalous or cases that "prove" a more general point. Is the role of in-depth analysis to assess the value of preferred theories, to lead us to new propositions, and/or to gain better insights into cases deemed to be of intrinsic interest? The nested analysis approach provides a strong foundation for adjudicating among the competing goals and inferential logic associated with case selection strategies, asking the scholar to make decisions about case selection in the context of the assessment of the preliminary LNA, which includes an assessment of confidence in one's theoretical model.

When carrying out Mt-SNA, scholars should only select cases for further investigation that are *well predicted* by the best fitting statistical model. Recall here that a decision has already been made that cases outside the confidence interval are not of theoretical interest and should be treated as unexplained "noise." Country cases that are on, or close to, the 45-degree line (plotting actual dependent variable scores against regression-predicted scores) should be identified as possible candidates for in-depth analysis. As discussed previously, in this instance SNA provides a check for spurious correlation and can help to fine-tune a theoretical argument by elaborating causal mechanisms. Although intensive investigation of "on-the-line" cases may lead to the identification of alternative explanations, the primary goal is to assess the strength of a *particular* model. As such, there is little value to the pursuit of cases that are not well predicted by the model.

Moreover, when carrying out Mt-SNA, one should select cases based on the widest degree of variation on the *independent* or *explanatory* variables that are central to the model. Because the goal is to demonstrate the robustness of a particular causal argument, the onus on the scholar is to identify process-tracing evidence from cause to effect. The opposite approach—of starting with the outcome and working backwards—would be much less efficient given the assessment of confidence in the original model. By selecting cases with varied scores on the explanatory variables, the scholar can use the SNA to demonstrate the nature of the predicted causal effect associated with the model in contrasting contexts.

Both Swank (2002) and Martin (1992) provide examples of book-length studies in which early chapters report statistical analyses that pave the way for Mt-SNA. According to Martin (1992, 92), "For those variables that showed statistically significant effects, the analyses complement the case studies by improving our confidence in the generalizability of our results." In each case, LNA provides initial confirmations of the author's core hypotheses and dismisses several rival hypotheses. However, in both cases, the authors acknowledge that questions about causality arise and that a range of possible mechanisms could be linking independent and dependent variables. As a result, they both select cases based on different scores on the central hypothesized explanatory variables and demonstrate the plausibility of their hypotheses by tracing the impact of alternative scores on those variables to predicted outcomes in the respective cases. (Graphically, this would be akin to selecting cases such as B, D, E, and F from Figure 2, in which a range of predicted values are considered.) Both scholars are deliberate in this approach For instance, Martin (1992, 11) writes, "This quantitative work allows me to further refine these hypotheses and provide a framework for the case studies that follow." Both Swank and Martin report additional findings and nuances about the cases they describe beyond demonstrating the plausibility of hypothesized relationships from the statistical results. For example, Swank points out that large-scale variables such as "international capital mobility" (captured in the LNA) are connected to discrete policy outcomes such as social expenditure through specific historical *episodes* (presumably distinct from an argument in which the mechanism is through long-term trends, or slow shifts), such as German unification or Italian political system restructuring (Swank 2002, 278). Within the case studies, we observe how actors behave, and we are presented with a more transparent accounting of causal mechanisms. Compared to an otherwise quite similar study such as Garrett's (1998) examination of the role of partisan politics in mediating the pressures of globalization, which presents only statistical findings, Swank's uncovering of cases and mechanisms provides significant additional evidence and insight. In the absence of such SNA, we would have been left to imagine the multiple causal pathways possibly associated with the

FIGURE 2. Case Selection from a Hypothetical Regression Analysis



statistical associations, and with greater skepticism about the general robustness—that is, non-spuriousness—of the results.

On the other hand, a very different set of strategies for case selection should be adopted in the case of Mb-SNA. First, *at least one* case that has not been well predicted by the best-fitting statistical model should be selected. Although it may be useful to select additional cases that are on the best-fit line for comparative analysis, the assessment that the preliminary statistical model was not sufficiently robust or that there were not sufficient data available to test certain critical hypotheses compels the scholar to examine cases that are *not* explainable by the right-hand-side variables included in the preliminary LNA. It is important to keep in mind the distinction between cases that are not well explained by the model (say, more than 2 standard deviations from the predicted value) and truly extreme cases that are several standard deviations from any other cases (e.g., case "H" in Figure 2). In the latter instance, the extreme nature of the case placement makes it more likely that the outcome was produced by a different causal process than most of the other cases in the population (and/or that some measurement error was involved). When such a distribution of cases is presented, case selection will hinge on whether the scholar is more interested in "making sense" of that deviance, or of developing a general theory that directly accounts for greater numbers of (less extreme) country cases.

Only when the scholar has good reason to believe that a particular case is "on-the-line" for entirely spurious reasons would it be useful to select such a case for Mb-SNA. However, in such instances the heuristic value of the preliminary LNA becomes increasingly obscured, and hence, of limited value.

In contrast to the Mt-SNA, case selection in Mb-SNA involves selection of cases based on initial scores on the *dependent* variable. Because Mb-SNA proceeds with vaguer theoretical hunches, the central goal is to try to account for important patterns of variation on the outcome.[23] Although it is important to try to ensure that among the cases selected there is sufficient variation on the explanatory variables of greatest interest at the outset, this is of secondary concern because there is much less confidence at the outset of the SNA that such variables will be significant when the research and analysis are complete. The very nature of Mb-SNA implies that we may lack the scores on the explanatory variables of interest at the outset of the project, making it impossible to use the explanatory variables for case selection. Although the strategy of selecting on the dependent variable has been a potential pitfall for much small-N scholarship, the nested approach provides

[23] Certainly, much social science analysis begins with the question, "What is the effect of X?," but almost always, there is a clear Y or outcome in mind. Such instances are examples of "strong theory." It is very rare that a scholar will start with an explanatory variable, but search inductively for an outcome to explain.

important correctives: the preliminary LNA provides a framework for selecting cases that vary widely on the variables of interest, and to the extent that the scholar hopes to draw general conclusions about the application of the resulting model, nested analysis involves the assessment of the hypothesis in subsequent LNA (discussed in the following section). Because causal inference in the nested approach does not rely solely on the small-N portion, the standard pitfalls of selection bias are less likely to lead to faulty inferences.

Using nested analysis, the preliminary LNA can be used to motivate structured comparisons for SNA, including a mix of "on-the-line" and/or "off-the-line" cases. In the simplest manifestation, when countries that would ordinarily be predicted to have similar outcomes wind up with different outcomes, perhaps on either side of the regression line, and with at least one case outside the confidence interval, scholars are presented with useful analytic puzzles that merit further examination (e.g., cases A, D, and C in Figure 2). Along these lines, the use of the nested approach could help to expand the scope of structured focused comparisons. Although there is a long tradition of deploying variants of Mill's "method of difference" to gain analytic leverage in cross-national comparative analysis, the requirement of identifying similar cases tends to limit scholars to comparing cases within regions, forcing certain sets of comparisons to reemerge: "France/Germany," "U.S./Canada," "Brazil/Argentina," and so forth. To a large extent, the underlying logic of such comparisons requires that the scholar make the implausible argument that the two or more countries are "virtually identical" in every way except on the relevant independent and dependent variables. As typically practiced (i.e., in the absence of LNA), the method virtually precludes making comparisons of countries at different levels of economic development, because that factor is assumed to have a causal influence on most outcomes of interest to Political Scientists. For example, comparisons between the United States and India might ordinarily be dismissed as not particularly useful because of vast differences in levels of economic development. However, within a nested analysis, one might find in the preliminary LNA that indicators of development do not hold any explanatory weight for the outcome of interest, and that colonial legacy (Anglo in both cases) and state structure (federal in both cases) are important predictors of the outcome, leading to similar point estimates and compelling a focused analysis of the two countries. Alternatively, in a strategy that approximates Mill's method of agreement, one might select cases with *differing* regression predicted values, and attempt to explain similarities in outcomes (e.g., cases B and G). In either case, LNA can set the stage for a comparative analysis that might otherwise seem implausible. The juxtaposition of such country cases allows for the additional exploration of the role of rival hypotheses that might not have been considered in the LNA because of lack of theory or data.

The nested analysis approach provides a self-conscious strategy for what many case-oriented scholars already do in practice: begin a research problem with an intuition that a particular case defies conventional wisdom or theorizing about a particular phenomenon, and then proceed inductively to generate explanations and theories that account for that exceptionalism. When using nested analysis, a potentially important finding of the preliminary LNA is that variables ordinarily thought to be associated with the outcome turn out to be statistically unrelated in the large sample. Alternatively, if the preliminary LNA demonstrates that the case *was* well predicted by conventional variables, this would give good reason to rethink the intuition of the case's uniqueness. If the LNA confirms the case's outlier status, however, this provides strong justification for intensive study.

As an example of such a move, Coppedge (2005) motivates the question of patterns of regime change over-time in Venezuela through various engagements with theory and preliminary LNA.[24] On the one hand, he demonstrates that on its own, a variable measuring over-time changes in level of economic development does a relatively good job of predicting democratic breakdown in that case. On the other hand, the inclusion of other factors helps to provide a better fitting model of regime outcomes *more generally* (across a large sample of approximately 4,000 country-years), and such a model does *not* predict the observed over-time changes in Venezuela's political regime. From this perspective, the need for the case study is clear: existing wisdom on the subject could *not* account for an important political outcome, and there is room for a new hypothesis or set of hypotheses to help address this conundrum. To accomplish this, Coppedge engages in an inductive Mb-SNA. (Incidentally, it is important to note that when using pooled time-series cross-section data, the "country" is still the unit about which one tries to make inferences, but the inclusion of historical data implies an interest in accounting for dynamics or historical patterns that describe each country, in the context of time-varying parameters.)

Selecting Cases Randomly or Deliberately

Scholars using nested analysis also face choices about whether the selection of cases should be done randomly or deliberately (nonrandomly). Again, the best strategy depends largely on the goals of the SNA and also on scholarly tastes and the scholar's familiarity with

[24] My definition and label of the nested approach are somewhat different from Coppedge's (2005). He explains, "Nested induction consists of a case study nested within a large-sample quantitative analysis. This method has three steps: 1) explaining the case of interest as much as possible using large-sample empirical estimates of the impact of general explanatory factors; 2) using the large-N estimates to pinpoint what is not well explained by the general factors (the residuals), and 3) using traditional case-study methods to propose supplementary explanations for the residuals (1)." My approach is more expansive, incorporating a wider variety of research problems and results. Moreover, I opt for the label "nested analysis" instead of his "nested induction" because I see no reason to limit this form of research necessarily to inductive theory-building. Although case analysis almost inevitably demands induction, there is no reason that this approach could not be used to examine deductively derived propositions.

and access to certain case materials. In most cases, deliberate selection will be the most appropriate strategy, but there may be specific instances when, in the course of carrying out Mt-SNA, random case selection can be used to address specific concerns about investigator bias. In any event, explicit consideration of this option forces us to reflect on potential sources of bias and measurement error in SNA, which should be considered in all aspects of the nested analysis.

Though rarely used in practice, when carrying out Mt-SNA, it may be desirable and appropriate to use a *random* case selection strategy. In a work-in-progress, James Fearon and David Laitin (2005) elect to further test their statistical model (2003) with narrative analyses of a set of randomly selected cases.[25] Fearon and Laitin (2005) opt for this approach, arguing that the deliberate selection of cases risks high levels of investigator bias. In particular, they say that the random selection of cases can provide an opportunity for a "fresh reading from the standard literature about a country." Moreover, they are concerned that in-depth investigation of cases they know well will induce confirmation of theories based on the very information that was used to derive the theory in the first place. Importantly, the rationale for random selection is *not* the development of a *representative* sample, as is the case in other forms of research, including survey research. The number of cases involved is simply too small to generate a useful representation of the entire population of country cases.

There are strengths and weaknesses associated with the random case selection option. On the one hand, there is good reason to believe that this strategy *should* lead to less investigator bias—However, it is only appropriate when the model specified in the LNA provides a good fit and when the investigator is less interested in identifying new hypotheses than in assessing the degree to which the logic of the theory behind the statistical model actually resonates with causes and effects within particular case histories. If a scholar can actually apply a model to a country with which he or she had little initial familiarity, confirm the independent and dependent variable scores with new measures, and find theoretically predicted links between cause and effect, such findings would provide considerable confirmation of the robustness of the model. As Fearon and Laitin (2005) suggest, a good strategy is to stratify cases based on independent and dependent variable scores in order to ensure a wide range of variation in case scores while attempting to economize on the total number of case studies carried out.

Despite certain appeal in the reduction of bias associated with random selection, the promised benefits must be weighed against pragmatic investigator limitations. The very rationale of this strategy commits scholars to cases where they may lack the technical skills for careful readings of country data, and mostly, if not exclusively, to secondary sources that may already be heavily biased by a particular theoretical bent.[26] This strategy may be particularly problematic when scholars carry out research in issue areas for which a complete secondary literature does not exist (in the case of Fearon and Laitin (2005), their focus on civil wars implies that this concern does not hold), requiring scholars to probe deeply into primary materials in order to carry out the analysis. One solution would be to enlist country experts to collaborate on country-based research generated from random selection and to ask them to adjudicate among best-fitting models (while being blind to the preferred model). This is an ideal strategy from a methodological standpoint if such an opportunity is available and appeals to one's scholarly style, but it also imposes high research costs.

Although the random selection approach is an intriguing option, most scholars will likely opt for a deliberate, or nonrandom, approach to the selection of cases. Particularly because problems of selection bias do not apply in the LNA component of the nested analysis research design, minimization of this bias in the SNA component is not likely to justify the costs associated with random case selection. Indeed, as stated at the outset, many scholars are interested to see whether general theories can help to make sense of particular cases and do not view case analysis as merely a means for assessing general theories. When selecting cases deliberately, the standard benefits of SNA are much more likely to apply, including the ability of the scholar to gain access to (often highly heterogeneous) data and to sensitively analyze such data with an appropriate degree of contextual background to make valid comparisons across cases. For example, evidence of the harsh exchange of words in various legislative contexts is likely to have very different implications for how we interpret the degree of cohesion or polarization across polities, depending on the norms of parliamentary debate. Or, in the case of racial/ethnic politics, the "coding" of bigoted language and the subtle ways in which discriminatory practices get carried out may only be apparent to a seasoned analyst. Valuable field research, the quality of which is greatly enhanced through language skills, is more likely to be endeavored if country cases are deliberately selected.

Indeed, when engaged in Mb-SNA, random selection of cases should absolutely be avoided because such an approach would be tantamount to saying, "I don't have a good theory, and I don't have an intuition about why a particular case would be illuminating for constructing a theory," which is hardly a solid foundation for investigation. Of course, many scholars who find themselves engaged in Mb-SNA will arrive at this form of analysis because, as discussed in the previously cited examples, they had already identified cases of potential theoretical interest. Alternatively, such cases will be selected because a scholar believes that he or she has a particular expertise, such as language skills,

[25] However, they do not limit themselves to the selection of well-predicted cases.

[26] For example, see Lustick (1996) for a discussion of the problems of bias in secondary sources in political science research. For a more general discussion of the problem of random selection in SNA, see King, Keohane, and Verba (1994, 125–28).

background, or historical knowledge, or because of a particular interest in a case.

When a scholar is intent on studying a particular case or set of cases, the nested analysis approach obviates the need to make the artificial claim that the case is the best one for studying a particular research question. Rather, the approach allows the scholar to identify the particular information that he or she wants to glean from the in-depth analysis of almost any case, and then to assess the potential added value of such analysis relative to a larger body of theory and data.

Scholars engaged in Mt-SNA may also use deliberate case selection, but they should avoid using the specific case or cases that informed the initial development of the theoretical model (i.e., prior to the preliminary LNA) as the basis for *testing* the model. Such a constraint may be highly prohibitive for scholars with a wide-ranging knowledge of country cases, whose theorizing may be informed by several important cases. A next best solution would be to try to gather *new* information about the particular cases with which the analyst is more generally familiar and to attempt to "test" the LNA-verified hypotheses with such data. Alternatively, the analyst may deliberately select a case of substantive interest, but with little prior knowledge of case specifics, capturing most of the benefits of the random selection procedure.

ASSESSING THE FINDINGS OF THE SNA: THE NEED FOR FURTHER NESTING?

Moving between SNA and LNA, when taken to the extreme, could imply an endless loop of research, with the only end in sight being the intensive analysis of every country case. Clearly, this is not a helpful vision of the nested approach, both because it is impractical and because it is likely to violate social scientific preferences for parsimony (Gerring 2001, 106–7). There is always more to be learned, but it is necessary to establish a set of criteria and procedures to conclude the analysis, leaving unanswered questions to future research. Again, just as there are no absolute answers to such a question in the cases of LNA or SNA on their own, strict guidelines cannot be established for the nested analysis approach. Nonetheless, we can establish useful assessment criteria for making decisions about when to end the analysis. Contrary to Lijphart's original view of the possible interaction between different types of research methods in comparative political analysis—in which SNA was merely a "way station" for LNA[27]—in the nested analysis approach, LNA serves as a way station for SNA at least as often as the reverse. A fundamental interest in the understanding of specific country cases helps to anchor the analysis in the nested research design.

Two endpoints are clear: in the case of Mt-SNA, if one or more intensive case studies can demonstrate the validity of the theoretical model—which had already passed muster in the LNA—by plausibly linking cause to effect in the expected manner, then the nested analysis provides ringing support for the model (**End analysis I** in Figure 1). Although we do not know the exact sequence for how the analysis was actually carried out in these works, the Martin (1992) and Swank (2002) books appear to be examples of this route.

At the other extreme, in the case of Mb-SNA, if a coherent theoretical explanation for the outcomes cannot be formulated, this also implies a natural endpoint (**End analysis IV** in Figure 1). In this situation, neither LNA nor SNA could generate a robust finding, suggesting that either the research question was poorly formulated or the outcome is generated by a largely random process. This implies the project should be abandoned or substantially reconstituted to the extent that it would be recognized as a new project. In a discipline that tends not to value negative findings or atheoretical analyses, it should come as no surprise that there are no published examples of such a project.[28]

When the Model-Testing SNA Fits Poorly

In between these two extremes, as depicted in Figure 1, there are a series of assessments that must be made to establish the next steps and procedures for analysis. When engaged in Mt-SNA, if the analysis does *not* support the statistical model, the scholar must assess the reason(s) for this poor fit. As in social science more generally, assessments of the link between evidence and theory contain a subjective element, and scholars are likely to disagree about goodness of fit and the factors driving the distribution of the data. Although the nested analysis approach cannot completely resolve such debates, it specifies the parameters of the assessment and the steps that ought to follow particular conclusions drawn from the data and analysis.

Idiosyncratic Cases. On the one hand, the scholar may decide that the Mt-SNA did not support the model because the selected case was clearly idiosyncratic in at least one important way—that is, some extremely rare historical event or set of circumstances obfuscated the types of social and political processes that were in the original model, or the variable scores were incorrectly measured for some highly anomalous, case-specific reason. Moreover, the scholar may decide that such unique circumstances do not merit theoretical elaboration because the epiphenomenal sequence of events is unlikely to travel to other cases. In this instance, the scholar remains confident that the model estimated in the LNA is still a robust one and that the case selected for in-depth SNA was found to be "on the line," but not for the reasons justified in the model. Although it would be important to report the findings of the SNA in the analysis, the degree of emphasis on that narrative will be a question of scholarly tastes—specifically, a taste for highlighting typical cases versus puzzling or deviant cases. Nonetheless, if the poor fit is due to factors not

[27] As analyzed by Collier (1991, 13).

[28] Often referred to as the "file drawer" problem, in which our exposure to the full range of evidence is constrained by vast quantities of unpublished, and therefore inaccessible, negative results.

likely to be found in the larger sample, an additional case or set of cases should be selected for additional Mt-SNA. If additional analysis again fails to confirm the original statistical model, the scholar should become highly suspicious of the assessment of idiosyncrasy and consider that the model may not be accurately capturing the general process it purports to explain. In subsequent iterations the scholar might conclude that the SNA undermined the robust findings of the LNA.

Theoretical Flaws. Alternatively, the Mt-SNA may reveal important shortcomings in the initial model and/or the statistical results. In such an instance, the Mt-SNA would reveal that the statistical correlation was in some way spurious—the variables are not measuring what they purport to measure, or it becomes clear that the presumed causal order of the original model is not in evidence in actual case analyses, or other variables not identified in the LNA specification are observed to be doing the causal work. For example, suppose an initial theoretical model claims that presidential systems of government lead to personalistic styles of politics, and this is somehow confirmed by the LNA. If the Mt-SNA shows clearly that cases of presidentialism tended to have personalistic political styles even prior to the introduction of democratic politics, we would have good reason to challenge the original model. What started as Mt-SNA would need to become Mb-SNA. Additional inductive exploration, combined with appeal to a broader set of theoretical propositions, is clearly necessary.

When the Mb-SNA Suggests a New Model

Looking at the Mb-SNA side of Figure 1, an additional assessment is also required when the Mb-SNA generates a promising theoretical model. Having completed intensive study of one or a few country cases, the intensive case analysis component of the nested analysis is complete, and the only remaining assessment to make is whether the model can generate testable propositions through additional LNA.

On the one hand, if the new model relies on explanatory variables that are difficult to measure across many cases (e.g., complex cultural, institutional, or historical variables), it may not be possible to develop quantifiable indicators or a statistical estimator that captures the theoretical relationships. Or, a scholar may decide that he or she has uncovered an important theoretical anomaly that is worth explaining, but for which further LNA would provide no added value because no additional cases would score in the same way, meaning no further testing of the hypothesis could be carried out. Finally, a scholar may decide that the purpose of his or her scholarship is to use theory to understand the puzzle of a case, rather than the reverse. In any of these instances, the scholar can report the findings of the preliminary LNA and end with the SNA. In the case of Coppedge's (2005) study of democratic breakdown in Venezuela, this is clearly the path that is chosen, reflected in **endpoint III**. Coppedge is able to *explain* the specific outcome in Venezuela by highlighting that other theoretical models cannot account for the specific deviations of this case and by identifying a unique set of causal conditions that are not captured in other theoretic accounts. He leaves it for future research to determine whether the features identified as deterministic in Venezuela can be integrated into a more general theoretical model.

Alternatively, if it *is* possible and desirable to develop measures of the new variable(s) and to deploy reasonable statistical tests of the model, then a Mt-LNA is clearly in order. Not surprisingly, the findings from SNA can form the basis for valid LNA. Close-range analysis of one or a few cases can be akin to developing a survey instrument through open-ended interviews and focus groups using a small sample of cases before fielding a large-scale survey. That is to say that a scholar can evaluate and/or develop indicators to be used for the measurement of a large number of cases through close-range measurement of one or a few cases.[29] The scholar may build on the rectangular dataset used for the preliminary LNA and add variables or create new measures for old variables. Depending on the new insights derived about the scope conditions for the model—that is, the bounding of the population of cases to which the model ought to apply—the scholar may add cases and/or remove cases from the LNA. The SNA may suggest important, theoretically informed control variables and interaction effects when close-range study highlights the implausibility of a simple linear model applying across all country cases. Finally, the scholar may test new model specifications derived from the SNA within the LNA.[30] Regardless of the findings, the completion of this LNA should be reported, ending the nested analysis at **endpoint II**.

An excellent example of the move from SNA to LNA is presented in Lynch's (2002) study of the age-orientation of the welfare state in the advanced industrialized countries. She derives a set of hypotheses about why some countries seem to favor older citizens through intensive study of three policy areas in Italy and the Netherlands. These are further explored in a pooled time-series cross-sectional (TSCS) analysis of social spending in 20 Organization for Economic Cooperation and Development (OECD) countries between 1960 and 1996. She is able to address the conventional wisdom generated from the welfare-state literature, ruling out several key rival hypotheses, from both a cross-sectional and a longitudinal perspective—though she points out that there are some heroic assumptions involved in the analysis of cross-national TSCS data. The statistical analysis also confirms the relationships between her independent variables (program structure and mode of political competition) and an expenditure measure of her dependent variable. Unlike other

[29] For discussions of the relationship between alternative measurement approaches and issues of measurement validity, see Coppedge (1999) and Adcock and Collier (2001).

[30] Scholars should report findings based on the entire sample as well as on the sample with the cases from the SNA removed from the sample in order to assess the degree to which the cases that were used to build the new model may be driving the results in the Mt-LNA.

studies in which LNA preceded SNA, Lynch's study is a clear example of SNA driving hypotheses and statistical tests for the LNA. Indeed, it is much easier to interpret the results of the LNA having read the intensive case analyses because one can understand how the results reflect on the machinations of politics and policy outcomes in the two countries of interest. In particular, Lynch's arguments about the central determinants of policy development were motivated by close-range study, and it is hard to imagine that such hypotheses would have been generated in the absence of such analysis. LNA allowed her to examine the extent to which such findings were unique to her initial two cases, or relevant to a wider group of countries.

As another example, Martin (1992) considers a new set of regression analyses after presenting her case study of the Falkland Islands conflict. She realizes that a potentially unique factor—the impact of military involvement on sanctions cooperation—needs to be explored. Having been convinced that military involvement affected this particular case, she moves back to the LNA, but finds that military involvement had only a negligible effect on sanctions in the larger sample (153–6). In this way, SNA helps to *motivate* the exploration of rival explanations within particular cases, and more generally. Similarly, in his comparative study of the politics of taxation, Lieberman (2003) raises the possibility that Brazil's Catholic heritage had been a determining factor in the development of a tax state that was very different from South Africa's. Although the SNA uncovered no plausible evidence linking Brazilian taxation to this legacy, this rival hypothesis could be dismissed with additional confidence through further LNA which provided no statistical support for the alternative hypothesis.

Again, it is important to emphasize that the nested analysis approach presumes interest in positive and negative findings, and in the analysis of general patterns and of specific cases. If the Mt-LNA is robust, we have arrived at findings quite similar to those of endpoint I: two sets of empirical analyses confirm the validity of the results and the scholar can feel extremely confident in the general applicability of those results. If the Mt-LNA is not robust—if the new variables do not predict what we had hypothesized or if the larger model falls apart, then the scholar is left to explain why those results might not have applied in the LNA. It is up to the scholar to account for the more limited scope of the explanation, which needs to be understood in the context of the larger population of cases. Future research projects may be used to develop models with more general applicability, but in this instance, the scholar should report what has been discovered through the nested analysis.

CONCLUSION

Despite the constraints of a relatively small, finite, complex, and heterogeneous universe of cases for analysis, scholars continue to be interested in questions about the causes and consequences of patterns of politics at the country level of analysis. To date, existing strategies of large-N cross-country regression analysis, as well as small-N case study and comparative analysis, have each been found wanting. This article argues for a combined approach. Some scholars may already be practicing a variant of this approach in their analysis, but to a large extent, the steps involved in the analysis are not being fully reported. With the provision of a more complete specification of how this approach can work, scholars should find it easier to use the approach in a self-conscious manner and to provide a more transparent accounting of their research. This will facilitate evaluation and replication of results, provide greater analytic clarity (by demonstrating how various analytic results relate to one another), and provide a recognizable bridge between research traditions that often remain quite isolated from one another.

Nested analysis provides a stronger basis for causal inference than the sum of its small-N and large-N parts. Rather than emphasizing the *common* inferential logic of qualitative and quantitative research strategies—which is the hallmark of King, Keohane, and Verba's (1994) influential treatment of research methodology—the nested analysis approach emphasizes the complementary distinctiveness in these two modes of analysis and strategies for causal inference. The use of the mixed strategy helps to overcome potential sources of bias and to sort out spurious findings that might be produced in either SNA or LNA when carried out in isolation. The approach is particularly well suited to cross-national analysis, where investigators tend to be interested not only in general patterns (as one might be in the study of, say, voting behavior) but also in the analysis of specific country cases.

There are clearly real and perceived costs of integrating LNA and SNA. Perhaps most importantly, this seems to imply a substantial addition of work. Is this really two projects in one? Undoubtedly, more investigator effort is required than if the individual SNA or LNA components were used in isolation, because nested analysis demands multiple forms of measurement and causal inference, but it does not entail a simple addition of effort. Rather, by highlighting the specific utilities of each analytic strategy, the approach lightens the inferential burden that would ordinarily be carried by SNA or LNA when performed on their own. Moreover, the advent of the Internet and an accumulation of research continue to expand the scope of available datasets that may be usable. For example, in the area of democratization research, the Freedom House, Polity, and a host of other datasets provide time-varying indicators across a large number of countries. For students of the political economy of development, the World Bank, the OECD, and the International Monetary Fund publish time-varying economic and other data across most countries for several decades. Similarly, not all "in-depth" case analyses involve multiple years of field research. As Fearon and Laitin (2005), Swank (2002), Adserà and Boix (2002), and Reiter and Stam (2002) demonstrate, at least some of the benefits of SNA can be captured using readily available data sources without extensive primary research. Again, increasing access to a range of primary and secondary sources through the

Internet has made the research and analysis of cases and structured comparisons far easier than was the case for previous generations of scholars. Given constraints on the particular skills of any single investigator, the nested approach is well suited to collaboration. Although it is certainly possible that this approach could simply be used as a model for a more general dynamic of the research cycle, the particular strategies and tactics outlined here assume the combining of strategies and require clear consistency in the use of concepts and measures, which are often lost when different scholars respond to prior iterations of the same question.

Nested analysis is a pragmatic and methodologically defensible scheme for comparative analysis. In this article, I have detailed its potential benefits, not by merely accepting the compatibility of qualitative and quantitative modes of analysis, but by demonstrating how each can be used to inform the execution and interpretation of the other.

REFERENCES

Achen, Christopher H. 1982. *Interpreting and Using Regression*. Beverly Hills, CA: Sage.

Achen, Christopher H., and Duncan Snidal. 1989. "Rational Deterrence Theory and Comparative Case Studies." *World Politics* 41 (January): 144–69.

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95 (3): 529–47.

Adserà, Alícia, and Carles Boix. 2002. "Trade, Democracy and the Size of the Public Sector: The Political Underpinnings of Openness." *International Organization* 56 (Spring): 229–62.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. 1998. *Analytic Narratives*. Princeton, NJ: Princeton University Press.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to do (and not to do) with Time-Series Cross-Section Data." *American Political Science Review* 89 (September): 634–47.

Bennett, Andrew. 2002. "Where the Model Frequently Meets the Road: Combining Statistical, Formal, and Case Study Methods." Presented at the Annual Meeting of the American Political Science Association, Boston.

Brady, Henry E., and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.

Campbell, Donald T., and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.

Collier, David. 1991. "The Comparative Method: Two Decades of Change." In *Comparative Political Dynamics*. ed. D. A. Rustow and K. P. Erickson. New York: Harper Collins, 7–31.

Collier, David. 1999. "Building a Disciplined, Rigorous Center in Comparative Politics." *APSA Comparative Politics Section Newsletter* 10 (Summer): 1–2, 4.

Collier, David, and James Mahoney. 1996. "Insights and Pitfalls." *World Politics* 49 (October): 56–92.

Collier, David, Henry E. Brady, and Jason Seawright. 2004. "Source of Leverage in Causal Inference: Toward an Alternative View of Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. H. E. Brady and D. Collier. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.

Coppedge, Michael. 1999. "Thickening Thin Concepts and Theories." *Comparative Politics* 31 (July): 465–77.

Coppedge, Michael. 2005. "Explaining Democratic Deterioration in Venezuela Through Nested Induction." In *The Third Wave of Democratization in Latin America*, ed. F. Hagopian & S. Mainwaring. Cambridge, UK; New York: Cambridge University Press.

Fearon, James, and David Laitin. 2005. "Civil War Narratives." Theory and Research in Comparative Social Analysis Working Paper Series, 27. Department of Sociology, UCLA. http://repositories.cdlib.org/uclasoc/trcsa/27.

Fearon, James, and David Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97 (February): 75–90.

Garrett, Geoffrey. 1998. *Partisan Politics in the Global Economy, Cambridge Studies in Comparative Politics*. Cambridge, UK; New York: Cambridge University Press.

Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis* (Spring): 131–50.

George, Alexander L. 1979. Case Studies and Theory Development: The Method of Structured, Focused Comparison. In *Diplomacy: New Approaches in History, Theory and Policy*. ed. P. G. Lauren. New York: The Free Press.

George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.

Gerring, John. 2001. *Social Science Methodology: A Criterial Framework*. Cambridge, UK; New York: Cambridge University Press.

Gerring, John. 2004. "What Is a Case Study and What Is It Good For?" *American Political Science Review* 98 (May): 341–54.

Gerring, John. 2005. "Causation: A Unified Framework for the Social Sciences." *Journal of Theoretical Politics* 17 (April): 163–98.

Huth, Paul K. 1996. *Standing Your Ground: Territorial Disputes and International Conflict*. Ann Arbor: Univeristy of Michigan Press.

Jackman, Robert W. 1985. "Cross-National Statistical Research and the Study of Comparative Politics." *American Journal of Political Science* 29 (February): 161–82.

Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press.

King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Laitin, David. 2002. "Comparative Politics: The State of the Subdiscipline." In *Political Science: The State of the Discipline*. ed. I. Katznelson and H. V. Milner. New York; Washington, DC; Norton: American Political Science Association.

Lieberman, Evan. 2001a. "Causal Inference in Historical Institutional Analysis: A Specification of Periodization Strategies." *Comparative Political Studies* 34 (9): 1011–35.

Lieberman, Evan. 2001b. "Taxation Data as Indicators of State-Society Relations: Possibilities and Pitfalls in Cross-National Research." *Studies in Comparative International Development* 36 (January): 89–115.

Lieberman, Evan S. 2003. *Race and Regionalism in the Politics of Taxation in Brazil and South Africa, Cambridge Studies in Comparative Politics*. Cambridge, UK; New York: Cambridge University Press.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65 (September): 682–93.

Lustick, Ian S. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review* 90 (September): 605–18.

Lynch, Julia. 2002. "The Age of Welfare: Citizens, Clients, and Generations in the Development of the Welfare State." Ph.D. diss. University of California, Berkeley.

Mahoney, James, and Dietrich Rueschemeyer. 2003. *Comparative Historical Analysis in the Social Sciences, Cambridge Studies in Comparative Politics*. Cambridge, UK; New York: Cambridge University Press.

Martin, Lisa L. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton, NJ: Princeton University Press.

McKeown, Timothy. 2004. "Case Studies and the Statistical Worldview." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. ed. H. E. Brady and D. Collier. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.

Munck, Gerardo L. 1998. "Canons of Research Design in Qualitative Research." *Studies in Comparative International Development* 33 (Fall): 18–45.

Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.
Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
Reiter, Dan, and Allan C. Stam. 2002. *Democracies at War*. Princeton, NJ: Princeton University Press.
Rogowski, Ronald. 1995. "The Role of Theory and Anomaly in Social-Scientific Research." *American Political Science Review* 89 (June): 467–70.
Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *The American Political Science Review* 64 (December): 1033–53.
Swank, Duane. 2002. *Global Capital, Political Institutions, and Policy Change in Developed Welfare States, Cambridge Studies in Comparative Politics*. Cambridge; New York: Cambridge University Press.
Tarrow, Sidney. 1995. "Bridging the Quantitative-Qualitative Divide in Political Science." *American Political Science Review* 89 (June): 471–4.
Western, Bruce, and Simon Jackman. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88 (June): 412–23.
Zuckerman, Alan S. 1997. "Reformulating Explanatory Standards and Advancing Theory in Comparative Politics." In *Comparative Politics: Rationality, Culture, and Structure*. ed. M. I. Lichbach and A. S. Zuckerman. Cambridge; New York: Cambridge University Press.