Newsletter Staff

Hertie School of Governance

Berlin Germany

Editors

Mark Hallerberg
hallerberg@hertie-school.org

Mark Andreas Kayser
kayser@hertie-school.org

Assistant Editor

Linnea Kreibohm

Editorial Board

Klaus Brösamle

Cassandra Grafström

Anke Hassel

Stein Kuhnle

Piero Stanig

Contact
Hertie School of Governance
Friedrichstr. 180
10117 Berlin Germany
Tel: +49 30 259219-0
Fax: +49 (0)30 259219-111

Letter from the Editors
Medium N Methods

by Mark Hallerberg and Mark Kayser

Welcome to the second issue of the Comparative Politics Newsletter under our editorship. The theme for this issue is relevant and appropriate metholdogy when the number of observations is neither so small that a straight case study is the obvious approach but also not so large that one can speak of a "large n" approach.

Our first piece by Charles Ragin explains Qualitative Comparative Analysis (QCA), a method that has gained popularity especially in sociology but that has a growing influence in political science as well. He presents a simple case using Olav Stokke's (2007) research on the impact of "shaming" in international fishing regimes to illustrate the approach.

Simon Hug also discusses the application of QCA. He notes the effectiveness of the approach, but he also cautions that one should be careful about measurement error in the dependent variable. One should also not use the technique for inductive theoretizing.

Another approach to deal with the small n issue is "analytic narratives." Luz Marina Arias considers the state of the approach, which Robert Bates, Avner Grief, Margaret Levi, Jean-Laurenth Rosenthal and Barry Weingast first introduced in 1998. She emphasizes that the focus of the approach is on the development and testing of theory as well as on the identification of causal mechanisms. Analytic narratives enable scholars to evaluate the validity of an explanation in the context of small samples. They are particularly useful for identifying the enabling conditions for particular

events. They do not, however, provide universal laws of behavior.

Kevin Clarke and David Primo discuss the relationship between empirical and theoretical models. They argue that, while both types of models are important, one should not think of empirical models as testing theoretical models. From the perspective of the theme of this issue's theme, it is beside the point whether one has enough observations to test a given theoretical argument. Given that comparative politics has been a relative newcomer to the "model testing game," it is well positioned to be the first sub-field to abandon the approach.

Tom Clark and Drew Linzer focus on the appropriate use of fixed effects and random effects in empirical models. Fixed effects, such as countries or regions, are ubiquitous in comparative politics. Assuming a correctly specified model, the authors argue that fixed effects lead to unbiased estimates but potentially to greater sample to sample variability. If the unit effects are correlated with the fixed effects, a random effect model will introduce some bias, but it will also constrain the variance and lead to values that are closer to the true value. They then propose three rules of thumb scholars should use to decide which model is most appropriate for a given dataset. They also reject the use of a Hausman test to decide between fixed and random effects.

We also have our regular sections. In terms of comparative politics datasets, Jan Rovny discusses the 2010 iteration of the of the Chapel Hill Expert Surveys on Party Positions. This dataset covers most European Union countries as well as the main non-EU countries in Europe. An interesting finding to emerge from the work so far is that parties of both the right and

Section Officers

APSA-Comparative Politics

President

Kathleen Thelen
*Massachussetts Institute of
Technology*
kthelen@mit.edu

Vice-President/ President-Elect

Duane Swank
Marquette University
duane.swank@marquette.edu

Secretary/ Treasurer

Julia Lynch
University of Pennsylvania
jflynch@sas.upenn.edu

2013 Program Chairs

Dan Slater
University of Chicago
slater@uchicago.edu

Lily Tsai
MIT
l_tsai@mit.edu

Executive Council

Lisa Blaydes
Stanford University

Ernesto Calvo
University of Maryland

Erin Jenne
Central European University

Fabrice Lehoucq
*University of North Carolina,
Greensboro*

Staffan Lindberg
University of Gothenburg

Victoria Murillo
Columbia University

the left shifted further to the left on issues of reglation since the last survey was taken in 2006, and since the beginning of the global financial crisis.

Our last page provides a list of the distinguished award winners from this year. They will be formally announced at APSA 2013.

We regret that we do not have any submissions for one section of our newsletter, namely "Heard at the Confernce." We encourage readers to submit brief reports about panels they thought were interesting. The editors intended to provide reports themselves in this issue, but Hurricane Isaac unfortunately intervened and led to the cancellation of APSA 2012. Rather than wait for the conference, we would like to publish this newsletter, but please do consider submissions not only for the "Heard at the Conference" section but for the section on new datasets as well.

Mark Hallerberg is Professor of Public Management and Political Economy at the Hertie School of Governance.
hallerberg@hertie-school.org

Mark Kayser is Professor of Applied Methods and Comparative Politics at the Hertie School of Governance.
kayser@hertie-school.org

Table of Contents

Medium N Methods

Counterfactual Cases and Configurational Analysis

by Charles Ragin

Most case-oriented research is configurational—aspects of cases are interpreted and understood in the light of other case-specific aspects. Most variable-oriented research is not configurational, and case aspects are understood and interpreted in the light of broad cross-case patterns (e.g., via matrices of bivariate correlations). From a variable-oriented viewpoint, configurational analysis is overly ambitious, for it entails consideration of combinations of case aspects (i.e., the different ways case aspects may be configured), and the number of logically possible configurations of causally relevant conditions increases as an exponential function of the number of conditions.

An investigation with four causal conditions defines an analytic space with sixteen sectors; an investigation with eight conditions defines an analytic space with 256 sectors; and so on. Ideally, from the perspective of variable-oriented research, a researcher interested in configurations of conditions should examine empirical instances drawn from each sector of the analytic space defined by the causally relevant conditions. This type of analysis, however, is virtually impossible. Social phenomena are extraordinarily lumpy, and most cases typically reside in a very small handful of sectors. This observation holds for almost all nonexperimental research in the social sciences (i.e., almost all studies). When Ns are small or intermediate, the problem of limited diversity of cases is even more apparent because there are, inevitably, far fewer cases than analytic sectors.

Case-oriented researchers are familiar with this problem and address it, usually implicitly, via counterfactual analysis. Empirical cases are compared with hypothetical cases. For example, a researcher might ask whether England would have become as democratic as it did, as early as it did, without its revolutionary break with the past (i.e., without its Civil War; see Barrington Moore, Jr. 1966). This counterfactual question populates a sector of the analytic space with a hypothetical case, filling an empirical void. A more common (and more implicit) analytic step

> ## With Qualitative Comparative Analysis (QCA), it is possible to conduct counterfactual analysis using a procedure that mimics the practice of case-oriented researchers

in case-oriented research is to exclude from an explanation an observed condition that is inconsistent with theoretical and substantive knowledge. Suppose, for example, that historical research revealed that serfdom in England did not fade away nearly as early as scholars had previously believed. Would the perpetuation of serfdom appear as an important ingredient in the explanation of the early development of democratic institutions? Not likely.

The "easy" counterfactual in this example would be England without lingering serfdom, a counterfactual situation that would make the early appearance of democratic institutions, more likely, not less.

With Qualitative Comparative Analysis (QCA) it is possible to conduct counterfactual analysis using a procedure that mimics the practice of case-oriented researchers (see Ragin 2008, chapters 8 and 9). Based on the user's substantive and theoretical knowledge, different hypothetical cases are defined as either 'easy' or 'difficult' counterfactuals, depending on how they match up with empirical cases. Easy counterfactuals, in turn, can be used to simplify complex combinations of causally relevant conditions. Using QCA, the 'knowledge' that the researcher brings to the analysis can vary, and different assumptions about causal conditions can yield different 'intermediate' solutions to a given truth table. Still, these different solutions are all constrained to be supersets of the solution that defines all counterfactuals cases as 'false' (i.e., not leading to the outcome in question) and a subset of the solution that defines all counterfactuals cases as potentially 'true' (i.e., may lead to the outcome in question; see Ragin 2008). This essay illustrates the impact of using different assumptions in counterfactual analysis, using a data set with an intermediate N (eight configurations). Researchers who take advantage of QCA's procedures for counterfactual analysis routinely apply only one set of assumptions. As this essay demonstrates, however, different assumptions may be used to accomplish different analytic objectives.

Consider Olav Stokke's (2007) research on the impact of "shaming" in international fishing regimes. Countries enter into agreements regarding where, how, and how often to fish. These agreements are violated from time to time. After all, compliance is voluntary; explicit enforcement mechanisms do not exist.

The only immediate recourse is to shame the violators, exposing their infractions in the hope that they will change their behavior. The relevant domain of cases to investigate is defined by instances of shaming. Sometimes shaming works; sometimes it does not. The goal of the research is to determine the conditions linked to its success versus those linked to its failure.

A modified version of Stokke's data set on shaming is presented in table 1, which uses fuzzy sets instead of Stokke's original crisp sets. I use fuzzy sets here simply to illustrate that there is no need to limit analyses to crisp sets when the evidence is based on qualitative assessments. Using a four-value fuzzy set, for example, it is possible to code the condition 'shadow of the future' (whether the target of shaming needs to strike future deals with the fishing regime) as follows: 1 = yes definitely, 0.67 = likely, 0.33 = possibly, and 0 = definitely not. The outcome ('success') is coded according to six-value coding scheme: 1 = clearly compliant, 0.8 = mostly compliant, 0.6 = somewhat compliant, 0.4 = somewhat noncompliant, 0.2 mostly

noncompliant, 0 = clearly noncompliant. Even simple four-value fuzzy sets offer important analytic leverage beyond that offered by crisp sets.

QCA examines the evidence configurationally, with the goal of deriving the different combinations of conditions ('causal recipes') linked to the outcome. There are three main solutions: the complex solution, which defines cases in sectors not represented in the table as 'false' (i.e., unsuccessful shaming); the parsimonious solution, which defines these same hypothetical cases as potentially 'true' (depending on whether using them reduces complexity), and the intermediate solution, which defines hypothetical cases in these sectors as true only if they constitute 'easy' counterfactuals. An easy counterfactual is a combination of causally relevant conditions that does not exist, but according to theoretical and substantive knowledge is more likely to result in the outcome than an existing, empirical case of successful shaming. For illustration of the general idea of an 'easy' counterfactual consider the first row of table 1. This successful case is more "in" than "out" of the following sets: 'advice,' 'shadow,'

'inconvenient,' and 'reverberations;' and it is more out than in 'commitment.' One easy counterfactual would be this same array but with inconvenient switched to more out than in. If shaming succeeded when the behavioral change was inconvenient, it stands to reason that shaming also would have been successful, given this same array of conditions, if the behavioral change had been not inconvenient (i.e., more out than in the set 'inconvenient'). The knowledge-based assumption that defines this counterfactual as easy is that 'inconvenient' should be linked to intervention failure.

The assumptions that guide the initial counterfactual analysis of the evidence are that the presence of advice, commitment, shadow, and reverberations, and the absence of inconvenient should all be linked to successful shaming. Thus, assumptions regarding all five causal conditions shape the initial intermediate solution. The three solutions that follow from this coding are as follows:

Very little simplification has been ac-

Table 1: Fuzzy set representation of Stokke's evidence on "shaming" as an intervention

row #	advice	commit	shadow	inconven	reverb	success
1	0.67	0.33	1	0.67	1	1
2	1	0	0	1	0.33	0.2
3	0.67	0	0.33	1	1	0
4	0.33	0.33	0	0.67	0	0.4
5	1	1	1	1	0.67	0.6
6	1	0.67	0.67	1	0	0
7	0.67	1	1	0	0	0.8
8	1	0	0.33	0.33	0.33	1

Complex: success ≥ advice·shadow·inconvenient·reverberations +advice·~commitment·~shadow·~inconvenient·~reverberations + advice·commitment·shadow·~inconvenient·~reverberations

Parsimonious: success ≥ ~inconvenient + shadow·reverberations

Intermediate: success ≥ advice·~inconvenient + advice·shadow·reverberations

Note: The symbol '·' indicates combined conditions—set intersection, '+' indicates alternate combinations—set union, '~' indicate set negation or absence; and "≥ indicates a subset relation.

advice: whether the shamers can substantiate their criticism with reference to explicit recommendations of the regime's scientific advisory body;

commitment: whether the target behavior explicitly violates a conservation measure adopted by the regime's decision-making body;

shadow of the future: perceived need of the target of shaming to strike new deals under the regime; such beneficial deals are likely to be jeopardized if criticism is ignored;

inconvenience: the inconvenience (to the target of shaming) of the behavioral change that the shamers are trying to prompt;

reverberation: the domestic political costs to the target of shaming for not complying (i.e., for being scandalized as a culprit);

success: target of shaming responded positively.

complished in the complex solution, which is a common result when there are many vacant sectors in the multi-dimensional vector space defined by the causal conditions. By contrast, a great deal of reduction has occurred in the process of generating the parsimonious solution. This result is also common when there are many potential counterfactual cases, as in this example, and there is no distinction between easy and difficult counterfac-

between easy and difficult counterfactuals. In effect, all are treated as easy. Still, the parsimonious solution is straightforward in its interpretation: shaming is successful when it is not inconvenient for the target to change behavior or when there is a combination of 'shadow of the future' and 'domestic reverberations.' The intermediate solution reveals that the parsimonious solution omits a causal condition that is common across all incidents of successful shaming—the support of the regime's scientific advisory body. Basically, the intermediate solution shows that the derivation of the parsimonious solution incorporated difficult counterfactuals. A simple diagram can be used to summarize the results of the intermediate solution by branching at the point of inconvenient versus ~inconvenient:

$$
\begin{array}{l}
\sim inconvenient \\
\quad / \\
advice \\
\quad \backslash \\
inconvenient \cdot shadow \cdot reverberations
\end{array}
$$

Different assumptions generate different intermediate solutions, within the constraints provided by the parsimonious and complex solutions. Consider the following example. A researcher is interested in differentiating between the causal conditions linked to successful shaming where there are domestic reverberations for being shamed versus those where there are no such reverberations. This distinction might serve as a proxy for government openness, with 'no domestic reverberations' signaling less open governments. In essence, the goal is to keep either 'reverberations' or 'no reverberations' from being eliminated in the derivation of the intermediate solution, so that the different contexts can be distinguished. To prevent these unwanted simplifications, it is necessary simply to avoid specifying a directional connection for 'reverberations' even though substantive knowledge certainly indicates that behavioral changes should be linked to the presence of 'reverberations.'

The results of this analysis are as follows (the complex and parsimonious solutions are not affected by the new coding of 'reverberations'):

$$
\begin{array}{l}
Intermediate: \ success \geq \\
\quad advice \cdot \sim inconvenient \cdot \sim reverberations + \\
\quad advice \cdot shadow \cdot reverberations
\end{array}
$$

This solution is a subset of the initial intermediate solution and states plainly that in situations where domestic reverberations are not an issue, the behavioral change must be not inconvenient and also have the support of the scientific advisory body. By contrast, in situations where domestic reverberations are an issue, it takes the combination of shadow of the future and a supportive scientific advisory body and may be ~inconvenient or inconvenient. The contrasting diagram branches at the point of reverberations versus ~reverberations:

$$
\begin{array}{l}
reverberations \cdot shadow \\
\quad / \\
advice \\
\quad \backslash \\
\sim reverberations \cdot \sim inconvenient
\end{array}
$$

A primary focus of the examination of a uniform intervention (shaming) across the set of cases receiving the intervention is on the question of 'how' the intervention succeeds—under what conditions it has the desired impact. This examination focuses on the different contexts that support successful intervention. In this light, a more complex intermediate solution may be preferred over a less complex intermediate solution. The key concern here is the amount of interpretive guidance offered by the intermediate solution. In general, more guidance is preferred to less guidance, as long as the amount of complexity permitted by the analyst does not become a hindrance to interpretation. From the perspective of empirical social science, it is best

to avoid both extremes—simplistic generalizations, on the one hand, versus the perception that that every case is unique, on the other.

This essay demonstrates how counterfactual analysis coupled with set theoretic methods can be applied to the study contexts and conditions. Nuanced counterfactual analysis allows the crafting of 'intermediate' solutions that provide textured accounts of intervention success and failure. An implicit assumption of this approach is that it is important to avoid two extremes. Overly parsimonious accounts of intervention successes and failures offer little analytic leverage when it comes to understanding 'how' an outcome comes about. The other extreme, allowing too much complexity, often culminates in the prosaic observation that every case is unique.

Charles Ragin is Professor of Sociology at the University of California, Irvine and a Part-Time Professor of Political Science at the University of Southern Denmark

cragin@uci.edu

Limitations of QCA

by Simon Hug

Researchers interested in comparative politics find with increasing frequency scholarly work relying on what Charles Ragin (1987) has labelled Qualitative Comparative Analysis (QCA). This method, relying on boolean algebra, allows researchers to evaluate deterministic hypotheses based on possibly complex causal pathways. It has been used in various research areas (for extensive lists of such work, see Rihoux 2006, Yamasaki & Rihoux 2008) and both Ragin (2000) and other scholars have expanded the method to cover variables other than dichotomous ones.

While the considerable strength of this approach is that it forces comparative scholars to be very precise about their causal argument and the measurement of their variables, scholars also occasionally used this method for inappropriate purposes, as highlighted below. In addition, while for most methods of analysis, for instance quantitative ones, quite quickly textbooks informed users about the consequences of "violations in assumptions," this is a practice that has largely escaped QCA-scholars (for a notable exception, see Skaaning 2011). In what follows, drawing on Hug (2008), I briefly highlight two problems whose consequences, contrary to those in quantitative methods, have largely escaped scholars using QCA.

In its original formulation QCA Ragin (1987) envisions that a series of dichotomous independent variables (possibly in combination) form necessary or sufficient conditions (possibly many) for the presence (or absence) of a dichotomous outcome. As Ragin (1987, 45) emphatically argues, QCA allows scholars to evaluate deductively arrived at hypotheses of particular (and possible multiple) causal paths leading to the outcome (or its absence). As the method of evaluation of these hypotheses based on boolean algebra relies to a large extent on Mill's (1973), (1843), book 6 chapter 7 methods of agreement and of differ-

ence, critiques leveled by Mill (1973), (1843), book 6 chapter 7 himself or Bennett (2004) apply as well, with the notable exception that QCA allows for multiple causal paths.

As long as sufficient variation is present in our data (an issue raised by Bennett 2004) and we have no omitted variables (an issue discussed by Seawright 2005), using QCA deductively allows us, provided our variables are measured without error, to make valid causal inferences. If, however, measurement error in our variables is present, we may well be led to erroneous inferences. Carrying out a Monte Carlo analysis (MCA) on two quite prominent examples of QCA analyses (dealing with welfare systems (Ragin 2000, 292) and political mobilizations (Osa & Corduneanu-Huci 2003)) I can show that even small amounts of measurement errors in the dependent variable (a problem explicitly dealt

> *QCA is perfectly adequate to test deterministic hypotheses of (possibly complex) causal paths, provided there is no measurement error. If measurement error is likely (i.e., always) users of QCA are well advised to take this into account and employ tools that can inform us of the likelihood of biases.*

with by the workhorse of quantitative analysis, namely OLS) offsets QCA results considerably (Hug 2008). Few, if any QCA studies, however, assess the robustness of their results for such measurement errors (for two notable examples, see Stokke 2004, Ebbinghaus 2005).

While measurement error may be

rather directly addressed by relying on Dion's (1998), Braumoeller & Goertz's (2000) or Clark, Gilligan & Golder's (2006) approach to evaluate necessary conditions, which all address directly measurement error, the frequent use of QCA for inductive purposes raises the specter of an even more daunting problem. As QCA in Ragin's (1987) original design was not designed for such purposes, it is not surprising that in this inductive mode it either becomes a simple tool of data description with no inferential leverage (in King, Keohane & Verba's (1994, 7f) sense) or may mislead us considerably in the process of theory construction due to its strong sensitivity to the cases at hand. While a theory could delimit its scope, QCA applied inductively by definition cannot generate scope conditions. Using the same examples I show in my study (Hug 2008) that results from QCA analyses are very sensitive to the cases employed. Dropping randomly one or two cases from the studies evaluated I find again that the generated theories (and/or causal inferences) differ considerably.

What are the lessons of my MCA? QCA is perfectly adequate to test deterministic hypotheses of (possibly complex) causal paths, provided there is no measurement error. If measurement error is likely (i.e., always) users of QCA are well advised to take this into account and employ tools that can inform us of the likelihood of biases. For inductive purposes, QCA is inappropriate, because its conclusions are very sensitive to the set of cases used. In the absence of clear scope conditions, the QCA results can at best describe the data at hand. They cannot provide any guidance about causation.

Simon Hug is Professor at the Department of Political Science and International Relations, University of Geneva.

simon.hug@unige.ch

Analytic Narratives: A Solution to the Small n Problem?

by Luz Marina Arias

The "analytic narratives" methodology stands on a deliberate balance between context-specific detail and theoretical model building. As such, analytic narratives are capable of uncovering and evaluating explanations in the context of a limited universe of observations. The methodology emphasizes both explanation and empirical testing. A detailed contextual account acknowledges the uniqueness of particular situations and allows for a carefully developed conjecture, while the theorizing of causal mechanisms reveals the reasoning and assumptions behind the conjecture. This framework then facilitates testing by confronting the empirical evidence with changes in the model parameters and with competing alternative explanations.

Robert H. Bates, Avner Grief, Margaret Levi, Jean-Laurenth Rosenthal and Barry R. Weingast offer the first systematic outline of the key elements of analytic narratives in their collection Analytic Narratives (1998). Their goal is to "construct logically persuasive and empirically valid accounts that explain how and why events occurred" (1998, 13). The project makes explicit the methodology that scholars in politics and international relations adopt when combining historical and comparative research with logical rigor (e.g. Ferejohn 1991; Levy 1990-91; and Myerson 2004). Analytic narratives, however, underscore the mathematical corpus of rational choice theory over heuristic sketches.

The importance afforded to historical and institutional detail, however, sets analytic narratives apart from most rational choice approaches. Rational choice scholars commence from a general model and then test their hypothesis with appropriate data. In contrast, analytic narrativists formulate and refine the model itself in interplay with the context-specific institutional elements of the narrative. As such, analytic narratives are driven by specific cases and seek to explain particular events.

The first step in constructing an analytic narrative is acquiring in-depth knowledge of the context and the historical process of the historical phenomenon of interest. This detailed account is essential to isolate the relevant strategic elements in the interaction: the key actors, their goals, and the rules that structure their behavior. These elements can then be formalized in a model. The analytical framework specifies the choices, constraints, and trade-offs the actors face in the phenomenon in question. The outcomes predicted by the causal explanation are then confronted with the narrative; the narrative serves to assess the predictions and to arbitrate among possible explanations in instances of observational equivalence. Further refinement of the model, and collection of more historical detail, can result from additional iterations between analytics and history.

The theory provides categories and a framework that constrain the conjecture about the causal mechanism and the relevant counterfactuals. A narrative without analytical rigor is practically unconstrained in laying out explanations. Theory helps highlight the issues to be explored and the general considerations and evidence that need to be examined. Accordingly, explanations based on theory alone are inadequate. For instance, games can yield multiple equilibria. The empirical context provides a rich account of actions and circumstances that helps develop and complete the explanation. The predictions of the theory must follow deductively, but the model need not provide the bulk of the explanation. A well-confirmed causal claim about why and how certain outcome obtained can be accounted for mostly by the narrative.

Analytic narratives, however, require an awareness of the types of interactions that can benefit most from the methodology and of the appropriate choice of theoretical framework. Analytic narratives serve best to explain micro-level social phenomena and not the structural conditions under which social interaction takes place. Macro-level structural factors are taken as exogenous, which implies that changes in such factors need to be incorporated as moves by 'nature' and not treated in an analytic fashion. To the extent that the formalization relies on game theory, analytic narratives are constrained to causal explanation found in strategic interactions among individuals, or actors that can be regarded as such. The historical and institutional detail underpinning analytic narratives can and must provide justification for the behavioral assumptions; that is, the cognitive, coordinative, and informational abilities of individuals. Such detail also allows scholars to evaluate whether to modify the theory to incorporate uncertainty or use alternative theoretical frameworks relying on different behavioral assumptions, e.g. behavioral or evolutionary game theoretic models.

The identification of causal mechanisms and context-specific richness of analytic narratives enable scholars to evaluate the validity of an explanation in the context of small samples. Furthermore, because the causal mechanism reveals the reasoning and assumptions, and because the predictions follow deductively, explanations derived from analytic narratives may be applied to other settings. Nonetheless, analytic narratives identify the enabling conditions for particular events rather than seek universal laws of behavior. Even though there are important limitations to the generalizability of explanations and to the scope of phenomena that analytic narratives explain best, the methodology can provide testable explanations which, when handled with care, may be applied to other settings.

Luz Marina Arias is Research Fellow, Center for Advanced Study in the Social Sciences, Juan March Institute.   larias@march.es

The Modeling Enterprise in Comparative Politics

by Kevin A. Clarke and David M. Primo

Comparative politics, almost alone among the subdisciplines, has kept alive a robust tradition of both quantitative and qualitative research. In recent years, however, these traditions have moved significantly closer to one another. Comparativists of all stripes now commonly follow what Keith Krehbiel (1991, 15) describes as the "orthodox tenets of positive social science"; that is, deducing hypotheses from a verbal or formal theoretical model, testing them, and then drawing conclusions about the theoretical model. Geddes (2003, 87), for example, writes "Coherent deductive arguments can be devised to explain constituent processes, and hypotheses derived from the arguments can be tested." Not to be outdone, Bates (2008, 8-9) claims that adhering to these tenets sets his work apart:

> I proceed in a different fashion. I start by first capturing the logic that gives rise to political order. While I, too, test hypotheses about the origins of disorder, I derive these hypotheses from a theory. By adopting a more deductive approach, I depart from the work of my predecessors.

This approach to "testing" theoretical models is known as hypothetico-deductivism, and it suffers from a number of defects, particularly when applied to political science. Deductions work in a particular way. Truth flows down a deductive system (if the premises are true, then the deductions must be true), but it does not flow up a deductive system (if the deductions are true, the premises may or may not be true).

Now consider a theory in political science. If the theory is true, then the deductions drawn from it must be true. Testing is therefore irrelevant. If the theory is not true, then the deductions drawn from it may be true or may be false, and the logical connection between the theory and the deductions is broken. In this case, testing cannot tell us anything about the theoretical model. Either way, testing the deductive consequences of a theory tells us nothing informative about the theory itself. (Table 1 depicts these two possible states of the world.) Of course, in comparative politics, as in the rest of political science, we routinely use false assumptions in our theories. We make assumptions such as rationality that we know simplify a far more complex reality.

What then should we do? In our recent book, *A Model Discipline*, we draw on an analogy between models and maps first made by the philosopher Ronald Giere. Think about maps for a minute. Maps are characterized by limited accuracy, partiality, and purpose-relativity. Take a Boston subway map as an example (http://metro-underground-maps.blogspot.com/2012/05/boston-subway- map-mbta.html). The map has limited accuracy and is in many ways factually wrong; if you deduce a hypothesis from the map such as, "the neighborhood of Mattapan is south of the city of Braintree," and test it, you would discover that the opposite is true. The map is also partial; it displays some features of the area and not others. Finally, it is purpose-relative. The map is useful for riding the subway, but it is of little use for anything else. Anyone attempting to use the map for walking or driving around the city will become hopelessly lost.

The limited accuracy and partiality of the subway map does not mean that it is somehow false. Whether the map is true or false is the wrong question; a map is an object, and objects cannot be false (or true). The right question is whether the map is useful. Subway officials can evaluate the map, not through deductive testing, but simply by handing it to subway riders and asking if it helps them negotiate the subway. Despite its many inaccuracies, the Boston map is really quite useful.

Models are like maps in that models have limited accuracy and are partial. In truth, we are aware of few political scientists who would disagree on these points. We are forever told that theoretical and empirical models make use of assumptions that "simplify" reality and include only one or two features of the political landscape. Only somewhat more controversially, we claim that models are purpose-relative in the same way that maps are. We show that theoretical

Table 2: Possible States of the World

	Cases	
	(1)	(2)
Assumptions Predictions	True True	False True or False
Connection between model and truth of prediction	Logical necessity	None
Informativeness of data analysis for "truth" of model	Uninformative	Uninformative

models can be useful in one or more of four ways: as foundational models, organizational models, exploratory models, and predictive models. Empirical models can be used for prediction, measurement, and characterization. (We show that a fourth use of empirical models, theory testing, cannot be justified beyond the relative comparison of models.)

Again, much of this way of thinking is not particularly controversial. The implications for the practice of political science, however, are controversial. We have been bombarded in recent years with the claim that science consists of proposing a theory and testing it with data. Not only would that definition come as a surprise to many in the hard sciences, it does not comport with what we know about the nature of models. Theoretical models are not "tested" with data; theoretical models are "tested" with models of data (a category that includes qualitative data). Why should one limited accuracy, partial, and purpose-relative model "test" another limited accuracy, partial, and purpose-relative model? Theoretical models can be useful without being tested, and empirical models can be useful in roles other than testing. It is rarely *necessary* to include both kinds of models in a single paper.

Theoretical models can be useful without being tested, and empirical models can be useful in roles other than testing. It is rarely necessary to include both kinds of models in a single paper.

That being said, there is a justifiable way of linking theoretical and empirical models, and it has to do with the concept of explanation. Empirical models cannot provide autonomous explanations, while theoretical models can. Knowing that democracies do not fight one another or that higher spending is associated with the number of parties in a government (Bawn & Rosenbluth 2006) does not provide an explanation. Theoretical models provide the explanatory "bite" that empirical models lack (see Chapter 6 of A Model Discipline for a discussion of what constitutes an explanation). Although choosing between rival explanations is not always necessary—complex events often have more than one explanation—it can be done with the tools of comparative model testing (Clarke 2007). Here the question is

not whether an explanation is "true," but which of a set of explanations is strongest given the available evidence.

Our argument has been misconstrued at times, so it is important to be clear. We believe that theoretical work is important. We believe that empirical work is important. Our deviation from current orthodoxy lies in our insistence that empirical models are not useful for testing theoretical models and that a new justification for linking the two is required. Comparative politics has been relatively late to the "model testing" game, which makes it the subfield best positioned to abandon the pretense that theoretical models must be tested, and that empirical work serves no purpose except in the service of theoretical model testing. We hope that our work helps comparativists lead the way.

Kevin A. Clarke is Associate Professor of Political Science, University of Rochester. kevin.clarke@rochester.edu

David M. Primo is Associate Professor of Political Science and Business Administration, University of Rochester. david.primo@rochester.edu

Deciding Between Fixed and Random Effects

by Tom S. Clark and Drew A. Linzer

Comparative research often confronts data that are grouped into units. We may observe many elections during a given year, multiple survey respondents within countries, countries measured over multiple years, and so on. These features give rise to a number of well-known complications when fitting linear regression models, including the potential for both bias and inefficiency in estimates of the ef-

fects of our variables of interest (e.g., Greene 2008). One of the first decisions researchers must make in this situation is how to account for unit effects in their models—and in particular, whether to employ so-called fixed or random effects. Advice on this topic is plentiful (e.g., Robinson 1998, Kreft and DeLeeuw 1998, Greene 2008, Kennedy 2003, Frees 2004, Gelman 2005, Wilson and Butler 2007, Arceneaux and Nickerson 2009, Wooldridge 2010), but can also be confusing and contradictory (see Gelman and Hill 2007, 245, for a discussion).

In our working paper, Should I Use Fixed or Random Effects?, we conduct a series of simulation experiments that provide practical guidance to researchers debating the use of fixed versus random effects. Our results lay out a general set of conditions under which one or the other approach may be preferred. Both models have advantages, as well as disadvantages, to consider. Assuming a correctly specified model, the fixed effects estimator will produce unbiased coefficient estimates, but those estimates can be subject to high sample-to-sample variability. The random effects model

will—to the extent that the unit effects are correlated with the explanatory variables—introduce bias in coefficient estimates, but can also greatly constrain the variance of those estimates—leading to estimates that are closer, on average, to the true value in any particular sample.

To assess the overall quality of inferences under the fixed and random effects models, we compare the root mean square error (RMSE) of each estimator, for datasets of different sizes and with a range of characteristics typically encountered in comparative research. This provides a consistent standard by which to judge the bias-variance tradeoff. We further demonstrate that another common metric used to select between fixed and random effects— the Hausman (1978) specification test—is neither a necessary nor a sufficient condition for making the best choice.

Although fixed effects are often favored on the basis of their unbiasedness, we show that there can be costly increases in the variance of coefficient estimates in datasets where the independent variable exhibits little within-unit variation, or is sluggish over time. The fixed effects estimator performs especially poorly when the dataset contains very few units or observations per unit, or when the level of correlation between the independent variable and the unit effects is low. In these cases, the benefit of variance reduction provided by the random effects approach can easily outweigh the small amount of bias the model introduces.

In the alternative situation where variation in the explanatory variable is primarily within (rather than across) units, we find that there is rarely any substantive difference between the estimates produced by the random effects and the fixed effects models, as measured by the RMSE. Thus, the conventional view that any correlation between regressors and unit effects

The conventional view that any correlation between regressors and unit effects generates a level of bias that should disqualify the random effects model is unfounded.

generates a level of bias that should disqualify the random effects model is unfounded. Instead, the researcher may wish to consider other factors in the decision. For example, if one wants to make predictions about unobserved units (which is impossible using a fixed effects approach), then the random effects estimator can be safely employed. Similarly, if perfect (or near-perfect) collinearity between a regressor of interest and the unit effects precludes the use of a fixed effects estimator, one should not resist the random effects model as a useful alternative.

Why is the Hausman test not the most effective way to decide between fixed and random effects? The Hausman test is designed to detect violation of the random effects modeling assumption that the explanatory variables are orthogonal to the unit effects. If there is no correlation between the independent variable(s) and the unit effects, then coefficient estimates in the fixed effects model should be similar to those in the random effects model. But if the Hausman test does not indicate a significant difference, it does not necessarily follow that the random effects estimator is "safely" free from bias, and therefore to be preferred over the fixed effects estimator. In most applications, the true correlation between the covariates and unit effects is not exactly zero. Thus, if the Hausman test fails to reject the null hypothesis, it is most likely not because the true correlation is zero, but rather because

the test does not have sufficient statistical power to reliably detect departures from the null. When using the random effects model, there will still be bias (if perhaps negligible), even if the Hausman test cannot reject the null hypothesis. Of course, in many cases, a biased estimator (random effects) can be preferable to an unbiased estimator (fixed effects), if the former provides sufficient variance reduction over the latter, as just described. The Hausman test does not aid in evaluating this tradeoff.

Our analysis yields a series of general rules of thumb that should guide researchers when deciding how best to model their data. There are, in our view, three primary considerations: the extent to which variation in the explanatory variable is primarily within unit as opposed to across units, the amount of data one has (the number of units and observations per unit), and the goal of the modeling exercise. Finally, we note that however a researcher elects to proceed—using either fixed or random effects—one estimator or the other will always be superior to pooling the data into a regression model that ignores unit effects altogether. Fortunately, modern statistical software has made it just as easy to estimate the random effects model as it is to estimate the more traditional fixed effects model.

Tom Clark is Associate Professor of Political Science at Emory University. tom.clark@emory.edu

Drew Linzer is Assistant Professor of Political Science at Emory University.

dlinzer@emory.edu

Datasets in Comparative Politics

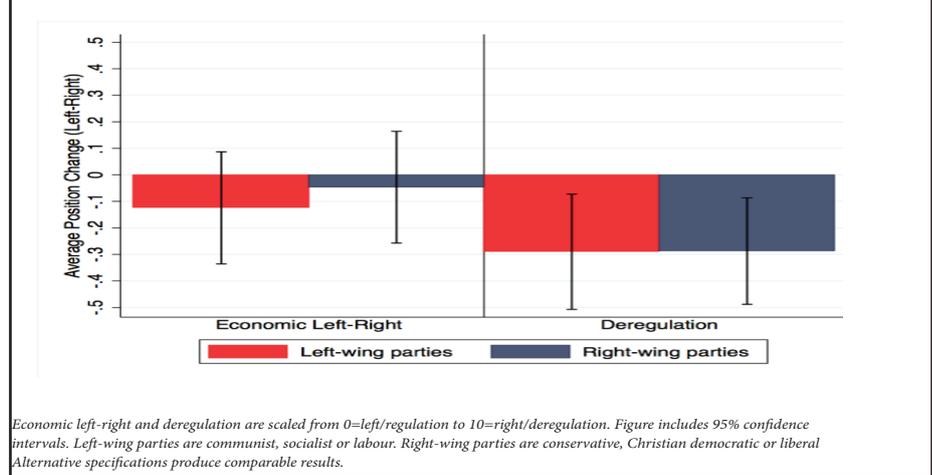## The 2010 Iteration of the Chapel Hill Surveys on Party Positions

by Jan Rovny

The Chapel Hill Expert Surveys (CHES) collect data on ideological and policy stances of national party leadership in Europe. It is an ongoing research project carried out by: Ryan Bakker, Catherine de Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen and Milada Anna Vachudova. The 2010 survey combines responses from 343 party specialists from all surveyed countries. This expert survey method allows researchers to obtain positions for a wide range of parties, in or out of parliament, in or between election years, and regardless of whether they publish a manifesto. The experts provide their assessments on the basis of broader knowledge, complementing information on what parties say, with analyses of what parties do.

The 2010 dataset covers 227 parties in 26 countries including all European Union member states, except Cyprus, Luxembourg and Malta, as well as two non-EU countries -- Norway and Switzerland. The dataset includes all political parties that obtain at least 3% of the vote in the national election immediately prior to the time of the survey, or that elect at least one representative to the national or European parliament. The 2010 data complements the previous CHES time-series with iterations in 1999, 2002 and 2006, which can be further connected to surveys dating back to 1984.

The survey is divided into three sections. The core of the survey asks about the general ideological orientation of parties on four dimensions: 1) general left-right, 2) economic left-right, 3) socio-cultural left-right,

Figure 1: Average Party Position Change Between 2006 and 2010

Economic left-right and deregulation are scaled from 0=left/regulation to 10=right/deregulation. Figure includes 95% confidence intervals. Left-wing parties are communist, socialist or labour. Right-wing parties are conservative, Christian democratic or liberal Alternative specifications produce comparable results.

4) European integration. The second section addresses more specific aspects of European integration, such as intra-party dissent over the EU, views on the single market, power of the European Parliament, and EU foreign and enlargement policies. The final section asks about party placement on and salience of general policy issues, such as taxation, immigration or the environment.

The 2010 CHES iteration adds an experiment evaluating the cross-national and cross-expert comparability of party placements on the economic left-right, socio-cultural left-right, and EU integration scales. After an expert evaluates the actual parties in her country of specialization, she is asked to place three fictional parties described by short vignettes on these three scales. Since these hypothetical parties are identical for all experts and countries, their placement effectively anchors each expert. Researchers can consequently assess the extent to which, say, Swedish expert placements tend to be shifted to the economic left compared with British equivalents.

Besides the ability to compare relative placements across countries and experts, the CHES dataset also allows researchers to assess the dimensional structure of political issues in different party systems. While providing party placements on general ideological di-

mensions, such as economic left-right, the data also contain party placements on specific policy issues, such as taxation, redistribution or deregulation.

Figure 1 demonstrates the utility of this by studying party position change in light of the economic crisis. It depicts the average placement change of left-wing and right-wing parties between 2006 and 2010 on the economic left-right dimension, and on the issue of deregulation. On the economic dimension, party position shifts are statistically insignificant. However, when looking at the more specific issue of deregulation, we can see that on average both, left- and right-wing parties, shift significantly to the left by about the same magnitude. While the economic crisis did not induce parties to shift their general economic views, it led them -- even those on the right -- to champion greater economic regulation. The data can be accessed at http://www.unc.edu/~hooghe/data_pp.php.

Jan Rovny is Research Fellow, Center for European Research, Dept of Political Science, University of Gothenburg.

jan.rovny@gu.se

Section Awards to have been announced APSA 2012

GREGORY LUEBBERT BOOK AWARD (Co-Winners)

Alan M. Jacobs, *University of British Columbia*
Governing for the Long Term: Democracy and the Politics of Investment.

Jeffrey A. Winters, *Northwestern University.*
Oligarchy.

LUEBBERT ARTICLE AWARD

Philip Roessler, *Duke University*

The Enemy Within. Personal Rule, Coups, and Civil War in Africa. *World Politics* 63:2 (2011).

SAGE PAPER AWARD

Rebecca Weitz-Shapiro, *Brown University*

"What Wins Votes: Why Some Politicians Opt Out of Clientelism."

LIJPHART/PRZEWORSKI/VERBA DATA SET AWARD

Project leaders: Ken Kollman, Allen Hicken, *University of Michigan*; Daniele Caramani, *University of St. Gallen*; David Backer, *University of Maryland*

Constituency-Level Elections Archive (CLEA)

POWELL GRADUATE MENTORING AWARD

(Named in honor of G. Bingham Powell, this new bi-annual award is for a political scientist who throughout his or her career has demonstrated a particularly outstanding commitment to the mentoring of graduate students in comparative politics.)

David Collier, *University of California at Berkeley*

About

The Organized Section in Comparative Politics is the largest organized section in the American Political Science Association (APSA). The Section organizes panels for the APSA's annual meetings; awards annual prizes for best paper, best article, best book, and best data set; and oversees and helps finance the publication of this newsletter, APSA-CP.

The section website is:

http://community.apsanet.org/comparativepolitics

Past newsletters can be accessed at:

http://community.apsanet.org/ComparativePolitics/ComparativePolitics-SectionNewsletter/

How to Subscribe

Subscription to the APSA-CP Newsletter is a benefit to members of the Organized Section in Comparative Politics of the American Political Science Association. To join the section, check the appropriate box when joining APSA or renewing your Association membership. You may join the APSA online at *https://www.apsanet.org/about/membform_start.cfm*

Dataset Review Submissions

If you have submissions for the dataset review section of the APSA-CP Newsletter, please email kayser@hertie-school.org .

American Political Science Association
1527 New Hampshire Ave, NW
Washington, DC 20036-1206
Phone: (202) 483-2512
Fax: (202) 483-2657
Email: apsa@apsanet.org

Heard at the Conference Submissions:

If you have submissions to the Heard at the Conference section of the APSA-CP Newsletter, please email hallerberg@hertie-school.org .

References

Counterfactual Cases and Configurational Analysis
by Charles Ragin

Moore, Barrington Jr. 1966. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World.* Boston: Beacon Press.

Ragin Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond.* Chicago: University of Chicago Press.

Stokke, Olav Schram. 2007. Qualitative Comparative Analysis, shaming, and international regime effectiveness. *Journal of Business Research* 60 (5): 501-511.

Limitations of QCA
by Simon Hug

Bennett, Andrew. 2004. Case Study Methods: Design Use and Comparative Advantages. In *Models, Numbers, and Cases. Methods for Studying International Relations*, eds. Detlef F. Sprinz & Yael Wolinsky-Nahmias. Ann Arbor: University of Michigan Press pp. 19–55.

Braumoeller, Bear F. & Gary D. Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science.* 44(4 October):844–858.

Clark, William Roberts, Michael J. Gilligan & Matt Golder. 2006. "A Simple Multivariate Test for Asymmetric Hypotheses." *Political Analysis.* 14(3):311–331.

Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." *Comparative Politics.* 30(2):127–145.

Ebbinghaus, Bernhard. 2005. "When Less is More: Selection Problems in Large- N and Small- N Cross-National Comparisons." *International Sociology.* 20(2):133–152.

Hanushek, Eric A. & John E. Jackson. 1977. *Statistical Methods for Social Scientists.* New York: Academic Press.

Hug, Simon. 2008. "An MCA of QCA." Département de science politique et relations internationales, Université de Genève.

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

Mill, John Stuart. 1973. Of the Four Methods of Experimental Inquiry. In A System of Logic, Ratiocinative and Inductive. *Collected Works of John Stuart Mill (Volume VII)*, ed. John Stuart Mill. Toronto: University of Toronto Press pp. 388–406.

Osa, Maryjane & Cristina Corduneanu-Huci. 2003. "Running Uphill: Political Opportunity in Non-Democracies." *Comparative Sociology.* 2(4):605–629.

Ragin, Charles. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantative Strategies.* Berkeley: University of California Press.

Ragin, Charles. 2000. *Fuzzy-Set Social Science.* Berkeley: University of California Press.

Rihoux, Benoît. 2006. "Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research." *International Sociology* 21(5):679–706.

Seawright, Jason. 2005. "Qualitative comparative analysis vis-a-vis regression." *Studies in Comparative International Development.* 40(1):3–26.

Skaaning, Svend-Erik. 2011. "Assessing the Robustness of Crisp-set and Fuzzy-set QCA Results." *Sociological Methods & Research.* 40:391–408.

Stokke, Olav Schram. 2004. Boolean Analysis, Mechanisms, and the Study of Regime Effectiveness. In *Regime Consequences: Methodological Challenges and Research Strategies*, ed. Arild Underdal & Oran R. Young. Dordrecht: Kluwer Academic Publishers pp. 87–120.

Yamasaki, Sakura & Benoît Rihoux. 2008. A commented review of applications. In *Configurational Comparative Methods: Qualitative ComparativeAnalysis (QCA) and Related Techniques*, ed. Benoît Rihoux & Charles C. Ragin. Thousand Oaks: Sage pp. 123–145.

Analytic Narratives: A Solution to the Small n Problem?
by Luz Marina Arias

Arias, Luz Marina. 2011. "Analytic Narratives: The Method," in *International Encyclopedia of Political Science*, ed. by Bertrand Badie, Dirk Berg-Schlosser, and Leonardo Morlino, SAGE Publications.

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

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R. Weingast (2000), "The Analytic Narrative Project," American Political Science Review, 94, 3, 696-702.

Ferejohn, John. 1991."Rationality and Interpretation: Parliamentary Elections in Early Stuart England." In *The Economic Approach to Politics*, ed. by K.R. Monroe, Harper Collins.

Greif, Avner. 2006. Institutions and the Path to the Modern Economy: Lessons from Medieval Trade, Cambridge University Press.

Kiser, Edgar and Howard T. Welser. 2007. "The Microfoundations of Analytic Narratives." *Sociologica* 3.

Levi, Margaret. 2003. "An Analytic Narrative Approach to Puzzles and Problems," in *Problems and Methods in the Study of Politics*, ed. by Ian Shapiro, Rogers Smith, and Tarek Masoud.

Levy, Jack S. 1990/91. "Preferences, Constraints, and Choices in *July 1914, International Security*, 15, 151-186.

Mongin, Philippe. 2010. "Analytic Narrative," in *The Encyclopedia of Political Science*, ed. by James E. Alt, Simone Chambers, Geoffrey Garrett, Margaret Levi and Paula D. McClain, CQ Press, SAGE Publications.

Myerson, Roger B. (2004), "Political Economics and the Weimar Disaster," *Journal of Institutional and Theoretical Economics*, 160, 187-209.

The Modeling Enterprise in Comparative Politics
by Kevin A. Clarke and David M. Primo

Bates, Robert H. 2008. *When Things Fell Apart: State Failure in Late Century Africa.* New York: Cambridge University Press.

Bawn, Kathleen, & Frances Rosenbluth. 2006. "Short versus Long Coalitions: Electoral Accountability and the Size of the Public Sector." *American Journal of Political Science.* 50 (2): 251-265.

Clarke, Kevin A. 2007. "The Necessity of Being Comparative: Theory Confirmation in Quantitative Political Science." *Comparative Political Studies.* 40 (July): 886-908.

Geddes, Barbara. 2003. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics.* Ann Arbor: The University of Michigan Press.

Krehbiel, Keith. 1991. *Information and Legislative Organization.* Ann Arbor, MI: University of Michigan Press.

References (Cont.)

**Deciding between Fixed and
Random Effects**
by Tom S. Clark and Drew A. Linzer

Arceneaux, Kevin and David W. Nickerson. 2009. "Modeling Certainty with Clustered Data: A Comparison of Methods." Political Analysis 17(2):177–190.

Frees, Edward W. 2004. Longitudinal and Panel Data: Analysis and Applications in the Social Sciences. New York: Cambridge University Press.

Gelman, Andrew. 2005. "Analysis of Variance—Why It is More Important than Ever." The Annals of Statistics 33(1):1–53.

Gelman, Andrew and Jennifer Hill. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press.

Greene, William H. 2008. Econometric Analysis, Sixth Edition. Upper Saddle River, NJ: Prentice Hall.

Hausman, Jerry A. 1978. "Specification tests in econometrics." *Econometrica.* 46:1251–1271.

Kennedy, Peter E. 2003. *A Guide to Econometrics.* Cambridge, MA: MIT Press.

Kreft, Ita G.G. and Jan de DeLeeuw. 1998. *Introducing Multilevel Modeling.* London: Sage.

Robinson, G.K. 1998. Variance Components. In *Encyclopedia of Biostatistics*, ed. Peter Armitage and Theodore Colton. Vol. 6 Wiley pp. 4713–4719.

Wilson, Sven E. and Daniel M. Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis.* 15(2):101–123.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd edition ed. Cambridge: MIT Press.

The 2010 Iteration of the Chapel Hill Expert Surveys on Party Positions
by Jan Rovny

No references.