Previous editions published by the American Political Science Association:
POLITICAL SCIENCE: The State of the Discipline (1983)
POLITICAL SCIENCE: The State of the Discipline II (1993)

Printed in the United States of America

The text of this book is composed in Electra LH
with the display set in Electra LH.
Composition by PennSet, Inc.
Manufacturing by Courier, Westford
Book design by Martin Lubin Graphic Design

Library of Congress Cataloging-in-Publication Data
Political science : the state of the discipline / ed., Ira Katznelson and Helen V. Milner.
p. cm.
Includes bibliographical references and index.
ISBN 0-393-05142-0
ISBN 0-393-97871-0 (pbk)
1. State, The. 2. Political science. I. Katznelson, Ira.
II. Milner, Helen V., 1958–
JC11.P65 2002
320—dc21 2002025547

W. W. Norton & Company, Inc., 500 Fifth Avenue, New York, N.Y. 10110
www.wwnorton.com

W. W. Norton & Company Ltd., Castle House, 75/76 Wells Street, London W1T 3QT

1 2 3 4 5 6 7 8 9 0

Political Science:

The State of the Discipline

Ira Katznelson and Helen V. Milner, editors
Columbia University

W. W. Norton & Company
New York | London

American Political Science Association
Washington, D.C.

CHARLES M. CAMERON AND
REBECCA MORTON

Formal Theory Meets Data[1]

■ | Political Methodology: Where Do We Stand?

We begin by providing the reader the context for this essay and that by Donald Green and Alan Gerber in this volume. Our device for doing so is a recent article by Nobel laureate James Heckman (2000) in which he discusses the current state of statistical methodology in our sister discipline, economics. Heckman's astute insights into econometric practice help illuminate the current role of statistical methods in political science, both by underscoring what is similar to economics and by revealing what is different.

CAUSAL ANALYSIS AND THE SEM APPROACH

As Heckman relates the recent history of econometrics, he assigns a central role to the linear simultaneous equations model (SEM), developed during the middle years of the twentieth century. The SEM allowed researchers to translate economic theory into well-posed, causally oriented empirical models. The concepts created by these researchers—exogenous variables, endogenous variables, causal effects, misspecification, omitted variable bias, the identification problem—after some delay entered political science, where they continue to provide the bread and butter of introductory methods training.

But, as Heckman tells the story, by the mid-1960s the SEM "was widely perceived to be an intellectual success but an empirical failure" (2000, 48). A simple example will help explain why. Consider the path

1. We thank Henry Brady, Alan Gerber, Don Green, Ira Katznelson, Gary King, Lisa Martin, Helen Milner, and participants in the 2000 Political Methodology Meetings, 2000 APSA meetings, and the "State of the Discipline" miniconference for very helpful comments on an earlier version. We also thank the more than 200 political scientists and economists who assisted in the literature survey that accompanies this chapter as a web page. Of course, the usual caveat applies.

Figure 1 | A Structural Model

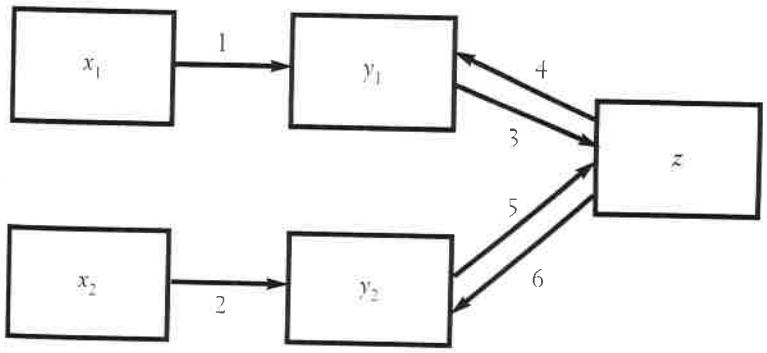


diagram shown in figure 1, which corresponds, to a three-equation model. The path diagram illustrates a structural model, detailing the causal relations between all the variables in the system; it shows how the "shin bone is connected to the knee bone, the knee bone is connected to the thigh bone," and so on. In the diagram, the variables x_1 and x_2 are external, or exogenous, because their values are determined outside the system; the variables y_1, y_2, and z are internal, or endogenous, because their values are determined within the system. Variables like y_1 and y_2 are sometimes called intermediate or intervening variables since their values are determined within the system but in turn determine the value of another variable z. The arrows labeled with numbers indicate the fundamental causal effects of one variable on another. For example, x_1 has a direct causal effect on y_1 via path 1, an indirect effect on z via paths 1 and 3, and an even more indirect effect on y_2 via paths 1, 3, and 6. Direct causal effects are sometimes called structural causal effects. Within the traditional SEM approach, all the structural causal effects in the figure are assumed to be linear, and the task of the empirical analyst is to use data to estimate each of them. Alternatively, if the interest were just variable z, the analyst might use the structural model to justify estimating a reduced form equation $z = z(x_1, x_2)$, relating z only to the two exogenous variables. This is a valid practice since the structural model shows that z ultimately is a function only of the external, or "forcing," variables.

Heckman suggests that the style of thinking illustrated in figure 1 was an intellectual triumph of twentieth-century social science. It's not hard to see why. The concept of stable causal effects is indispensable in most approaches to social science, and the structural approach offers a rigorous yet tractable way to think concretely about causal effects. Because the approach is so grounded in causal thinking, structural parameters are often social-scientifically transparent; they are readily interpretable in terms of theory. Hence, empirical estimates of them can be used to test theory. In

addition, empirical estimates of causal parameters can be used for forecasting or performing "what if" policy experiments. Finally, the SEM approach clarifies the limits of purely empirical evidence, by focusing attention on the necessity of identifying assumptions. It is for these reasons that Heckman calls the approach an intellectual success. Most political methodologists would agree.

Yet, as Heckman indicates, the SEM approach soon ran into empirical difficulties, at least in economics. To put the matter simply, only rarely did estimations uncover stable causal relationships (the direct path effects in figure 1). The problem was most notorious in empirical macroeconometrics but, as Heckman documents, widespread in other fields like empirical labor economics. Absent stable structural parameters, the framework illustrated in figure 1 implodes. Heckman suggests that most of what occurred in the methodology of empirical economics since the mid-1960s represents a response to the perceived failure of the SEM program to uncover stable causal parameters.

Post-SEM Developments in Political Science

Political scientists did not suffer the demoralizing failures of the empirical macroeconomists. Nonetheless, many of the new developments in political methodology have followed or even imitated the post-SEM moves in economics, though sometimes the intellectual origins in the SEM approach are overlooked.

Broadly speaking, there are four such moves. The first (and most conservative) locates the failure of the SEM approach in *tools*. Proponents of this approach argue that simple linear models were too rigid or otherwise inappropriate for political data. Their approach emphasizes the development and use of more appropriate, powerful, or flexible tools—duration models, event count models, and general additive models. It also emphasizes more powerful methods for estimation, including computer intensive methods like bootstrapping or Markov Chain Monte Carlo (MCMC) methods from Bayesian statistics. Not surprisingly, the tools approach is extremely popular among political methodologists. This volume does not contain a tools-oriented essay—they tend to be rather technical—but many are available for interested readers.[2]

The second and third approaches are somewhat more radical. The second locates the failure of the SEM approach in *theory*. It argues that the social scientific theories underlying early efforts were too weak or ill

2. Here is a selection of recent essays and monographs for different models: duration models—Box-Steffensmeier 1998, Gordon forthcoming, and Therneau and Grambsch 2000; count models—King 1989 and Cameron and Trivedi 1998; bootstrapping—Davison and Hinckley 1997; general additive models—Hastie and Tibshirani 1990 and Beck and Jackman 1998; and MCMCs—Congdon 2001, Gelman et al. 1995, and S. Jackman 2000.

formed and the link between the empirical analysis and the theories was too tenuous to sustain well-grounded analysis. We pursue this response in the remainder of this chapter. The third approach locates the failure in *data*. It argues that structural estimation with nonexperimental data is often doomed to failure. Consequently, it focuses on acquiring much better data—whether from laboratory experiments, field experiments, or so-called natural experiments. Stronger data can allow one to establish clear causal effects, often without deploying much social scientific theory. The essay by Green and Gerber in this volume pursues this line of thought.

The fourth approach is perhaps the most radical. It takes a step away from the causal thinking at the center of the structural approach. It argues that identifying assumptions making a strong distinction between exogenous and endogenous variables are untenable. Hence, the best one can do is to stay close descriptively to the data and make short-term forecasts. This line of thinking leads to vector autoregression (VAR) approaches in time series and neural net or other highly black-box approaches in cross-sectional data. The fourth approach is not represented in this volume, but interested readers may consult, for example, Beck, King, and Zeng 2000.

The first three approaches tend to have different adherents, who sometimes disagree intensely. But all three approaches are broadly complementary. Few of their adherents would argue against deep theory, closely tied to empirical analyses employing appropriate tools and strong, on-point data. In our view, the fourth approach runs the risk of throwing out the social scientific baby with the methodological bathwater, but data description and forecasting have a place too. In sum, political methodology is more heterogeneous and less naive than it once was. Yet almost all of its practitioners remain strongly committed to the ideal of *causal inference in service to causal reasoning*.

This Essay

The remainder of this essay explores new efforts to link theory and data in what we call (for want of a better term) formal empirical (FE) work. In addition, we have constructed a web page with supplementary material (a link may be found at http://www.columbia.edu/~cmc1). This web page lists hundreds of FE articles, books, and working papers, classified by topical subject in political science. The topics range across voting and elections, international political economy, war studies and international relations, legislative studies, executive studies, judicial politics, democratization, a grab bag of topics in comparative politics, and many, many more. Collectively, these studies give the lie to the claim that formal models are never tested in political science. This canard was an overstated but nonetheless plausible description of the state of the discipline in 1985 or even 1990. It grossly misrepresents the practice of political science ten years later.

The essay has the following organization. In the next section, we ex-

plore the empirical content of formal models, using a simplistic model as an expository device. The following section provides thumbnail sketches of some interesting real examples. The last section presents an incomplete and prejudiced overview of the substantive accomplishments of FE work.[3]

■ | The Empirical Content of Formal Models

When we say "social scientific theories underlying early efforts were too weak or ill formed, and the link between the empirical analysis and the theories too tenuous, to sustain well-grounded analysis," what do we mean? In this section, as a pedagogic device we construct a rather old-fashioned, tinker-toy formal model and show how to use it to structure empirical work. The model and empirical implementation illustrate in bare bones form the SEM approach, thereby providing a benchmark (or, perhaps, straw man). We then discuss why contemporary theorists view models of this kind as inadequate and how they are moving beyond them, and why some FE analysts see this style of empirical work as unsatisfactory and how they are attempting to forge stronger data-theory links. But we also discuss the virtues of inadequate models for structuring empirical work, for at some level, all models are inadequate.

A Motivating Example: Political Outcomes in a Democracy

What is the relationship between voters, interest groups, politicians, and political outcomes in a democracy? This is a central question in modern political science. Even a cursory summary of the many relevant literatures is beyond our scope. However, for purely pedagogic purposes, pluralist "theory" of the 1950s and 1960s affords a starting point. A caricature of pluralist notions is the parallelogram of forces: government policy reflects the vector of forces created by different pressure groups (Truman 1971 [1951]).

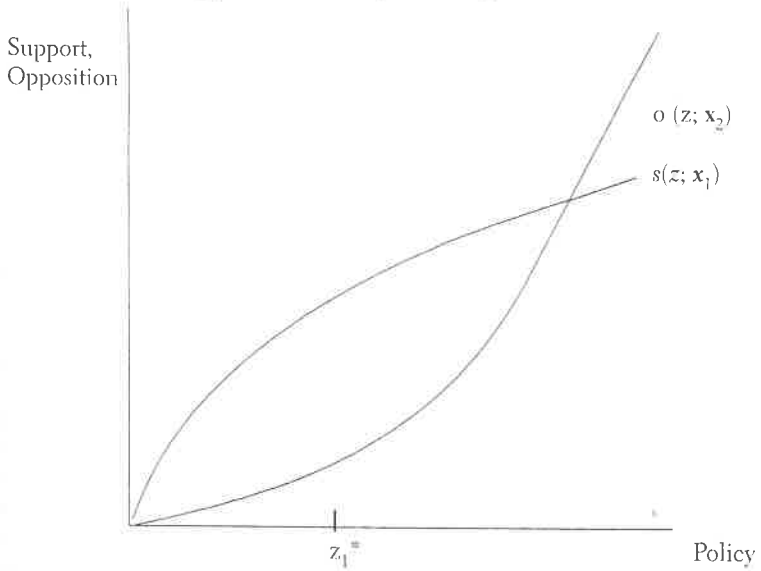
How might one formalize this parallelogram and use it to structure empirical work? The Chicago political economy models of pressure group politics represented an early effort to do so (Stigler 1971; Peltzman 1976; Becker 1983).[4] Most formal political theorists now see these models as inadequate. But they afford a relatively painless entree to FE work, because their extremely simple structure allows a clear demonstration of the SEM approach.

3. Morton's *Methods and Models* (1999) is a book-length exposition of many of the ideas in this chapter. Bates et al.'s *Analytic Narratives* (1998) provides a somewhat different take on FE work, stressing qualitative historical data.

4. Becker's version is rather different from Stigler's and Peltzman's. In particular, in his game theoretic formulation, interest groups are the actors. But both approaches, as well as more recent interest group models in political economy (e.g., Magee, Brock, and Young 1989) and rent-seeking models in public choice economics adopt a broadly similar, black-box approach to elections and political institutions.

In simplest and most schematic form, we imagine a single actor, the government, facing a political support function $s(z; \mathbf{x}_1)$ and a political opposition function $o(z; \mathbf{x}_2)$, where z represents a government policy (e.g., a tax, tariff, subsidy rate, or a liberalism-conservatism score for some complex policy). In addition, both support and opposition are functions of many other variables too ($\mathbf{x}_1$ and $\mathbf{x}_2$ respectively).[5] In the Chicago tradition, these functions are assumed to be everywhere twice continuously differentiable. In this sketch model, we imagine the government setting policy to maximize its net political support, that is, to maximize $n(z; \mathbf{x}_1, \mathbf{x}_2) = s(z; \mathbf{x}_1) - o(z; \mathbf{x}_2)$. Models in this tradition typically assume political support and opposition both increase with the level of the policy, but benefits do so at a decreasing rate while costs do so at an increasing one. Denoting the marginal change in s with respect to z by s' and the marginal change in that change with respect to z as s', and similarly for similar changes in o, it is as-

Figure 2 | Equilibrium in the Expository Example. z_1^* indicates the policy leading to maximum net support (that is, the greatest difference in "support" and "opposition").



5. The bold notation, e.g., $\mathbf{x}_1$, denotes a vector, that is, a group of variables. In contrast, x_1 denotes a single variable.

sumed that $s' > 0$, $o' > 0$, $s'' < 0$, and $o'' > 0$. Standard techniques from calculus show that the solution to the government's policy-setting problem is characterized by choosing a level of policy z^* that equates marginal political support and marginal political opposition (that is, $s' = o'$) provided $s'' < o'$ at z^* (which is true by assumption). This answer has an intuitive plausibility: if the government set policy lower than z^*, it could increase its net political support by increasing the level of the policy, while if it set it higher than z^* it could increase net political support by lowering the level of the policy. Figure 2 illustrates the solution graphically.

What is the empirical content of this tinker-toy model? On the one hand, it appears to have a great deal of content in the form of extremely strong and highly contestable assumptions. Among these: there is some entity that can be meaningful considered a policy-setting actor, and it seeks to maximize its net political support; there is something that can be meaningfully thought of as political support, which increases in the level of the policy; and so on. From this perspective, a test of the model could come through direct empirical evaluation of the basic assumptions. On some occasions, as discussed in Morton 1999, *assumption evaluation* of formal models is quite feasible, as when basic assumptions are amenable to straightforward empirical evaluation. When practical, assumption evaluation is appealing because of its directness and simplicity. Unfortunately, many assumptions are not amenable to simple empirical evaluation. Moreover, all theorizing involves abstract concepts and maintained assumptions. Direct assaults on these frequently bring the response "it all depends on how you think about it," or "you are being too literal," or even "you are missing the point." Further discussion then assumes an unproductive, quasi-theological quality. For these reasons, assumption evaluation only rarely proves decisive in practice.

An alternative approach concedes the model's maintained assumptions and then asks, *granting these*, what is the model's empirical content? In some sense this is a charity principle. But it is also affords an even tougher test than assumption evaluation, for if the model says nothing useful even *after* we grant it its underlying assumptions, then it has little to recommend it. Because this approach focuses on the model's empirically observable predictions, Morton (1999) calls this *prediction evaluation*.

If we grant the simplistic pluralist model its maintained assumptions, what is its empirical content in terms of empirically testable predictions? The answer is *nothing*, at least so far. Directly evaluating whether policy has been set to equate marginal support and marginal opposition is an impossible task. So the model has no real empirical content yet.

To imbue the model with empirical content, we must think harder about the impact of *observable* variables on other *observable* variables. Because the model is constructed to explain the level of policy—presumably an observable quantity—z is an obvious candidate for one variable. Observable exogenous variables (or observable but logically prior intervening ones) supply the other candidates. Hence, the question the model must answer is, what is the impact of (observable) exogenous variable x on endogenous variable z? More specifically, in this simplistic pressure group model, the logical chain of inference will run: effect of (observable) exogenous variable x on either (unobservable) intervening variable political opposition o or (unobservable) political support s; effect of (unobservable) support s or opposition o on (observable) policy level z; hence, effect of (observable) exogenous variable x on (observable) endogenous variable z.

There is a direct link between this reasoning and the logic of the SEM approach. In fact, the pluralism model has exactly the form shown in figure 1, with s corresponding to y_1 and o corresponding to y_2. A full-scale empirical implementation of the structural model would use data to estimate the direct effects on support and opposition of the exogenous variables and the policy level z, and the direct effects of support and opposition on effect z. However, *given* the theory embodied in the structural model, one is justified in moving to the reduced form policy equation $z^* = z(x_1, x_2)$. This is exactly the logic of the last part of the preceding paragraph.

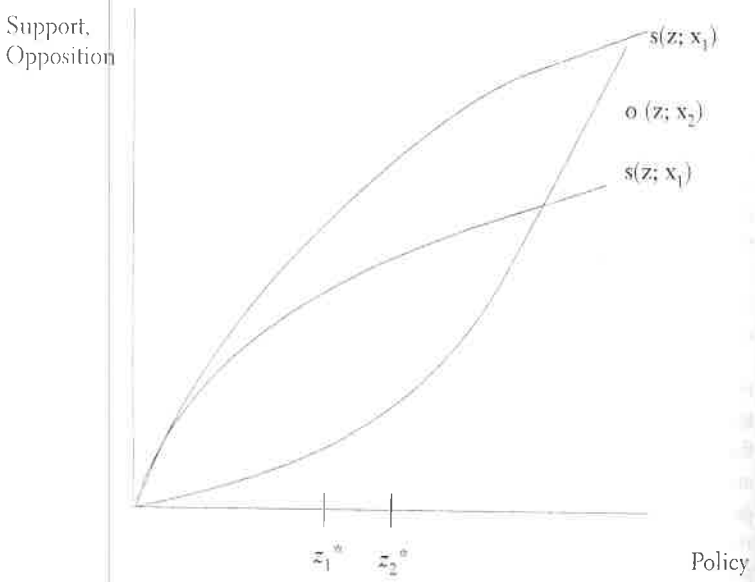
There are many plausible candidates for exogenous variables that enter the political support and opposition functions. Examples include the size of the groups that support or oppose the policy; the wealth of the members of those groups; the geographic dispersion of group members; the organization and procedural rules of the government, including the control of key proposal and veto points by supporters or opponents of the policy; the electoral rules that select politicians and allow supporters or opponents to reward or punish them for their actions; laws that control the use of money in politics; and so on. The silence of Chicago-style models on these points is their Achilles heel; momentarily, we discuss the consequences for theory and empirics of thinking hard about these matters. But to pursue the immediate pedagogic point, assume for the moment that somehow we specify that variable x_1 increases the government's political support at all positive levels of z while leaving political opposition unaffected.

Figure 3 illustrates the effect of x_1 on the support curve: as x_1 increases, the support curve rotates upward. In turn, this rotation changes the location of the point of maximum difference between the support and opposition curves. Consequently, z^* increases. Thus the theoretical result $\frac{dz^*}{dx_1} > 0$. (One reads this quotient as, "a marginal increase in x_1 strictly increases policy.") This hypothesis, and others like it, constitutes the empirical content of the model.

The figure illustrates the logic of the result but does not constitute a valid proof. A set of techniques called *comparative statics* supplies mathematical tools for carefully deriving such hypotheses from underlying assumptions.[6] In the interest of economy, we forgo a demonstration.

6. Explication of these tools may be found in most textbooks on mathematical economics and in many microeconomics texts. No political science textbook that we know exposits these techniques—a telling omission, perhaps.

Figure 3 | A Simple Comparative Static in the Expository Model. An increase in x_1 leads to a shift upward in the support curve, raising the level of the optimum policy from z_1^* to z_2^*.



However, it is important to note that the formal comparative static result $\frac{\partial z^*}{\partial x_1} > 0$ is not like the "hypotheses" discussed in statistics texts, which might better be described as empirical conjectures or, bluntly, hunches about data. Rather, within FE work, a hypothesis *is a formal mathematical result derived from the basic assumptions of the model.* In our view, the epistemological status of a formally derived hypothesis is quite distinct from an informal hunch about data, however acute.

How would one use the simplistic pressure model to structure empirical work? Working within the SEM tradition, first one would show that the reduced form policy equation $z^* = z(x_1, x_2, \ldots, x_m)$ is linear in the x's (some might simply assume this). Hence, the theory specifies $z^* = b_0 + b_1x_1 + b_2x_2 + \ldots + b_mx_m$. To turn this equation into a stochastic relationship that one could take to actual data, one would add a convenient white noise error term (say, u) to create $z = b_0 + b_1x_1 + b_2x_2 + \ldots + b_mx_m + u$. The error term would be rationalized as reflecting omitted variables, measurement error in z, or inherent randomness in society. Given a modest quantity of data, the parameters in this stochastic relationship can be estimated via multiple linear regressions, yielding actual numerical estimates, with standard errors, for each of the parameters, for example, $\hat{b}_1$ for b_1. It is important to note that the formally derived comparative static hypotheses correspond *exactly* to the parameters in the linear reduced form equation; for example, in the linear reduced form equation $\frac{\partial z^*}{\partial x_1} = b_1$. Consequently, the estimated coefficients, for example, $\hat{b}_1$, and their standard errors allow direct statistical tests of the model's formally derived empirical content!

The Virtues of Inadequate Models

Before examining the shortcomings of this style of formal model as theory and this methodology as a template for FE work, it is worth pausing briefly to consider the virtues of even inadequate formal models *as devices for structuring empirical work.*[7] These virtues are almost nascent in the simple parallelogram model, which after all is just a pedagogic device. Yet one can still discern their outlines. Three virtues stand out: clarity, rigor, and unity.

1. Testable hypotheses formally derived from explicit assumptions have the advantage of *clarity.* An overt casual mechanism generates the predictions in a transparent way. This transparent chain of logic may lead to surprises. It may uncover unexpected ambiguities. It may even reveal inadequacies in the underlying assumptions and provoke a reformulation of the theory. But in all cases, everything is laid on the table for inspection.

2. Testable hypotheses formally derived from explicit assumptions have the advantage of *rigor.* Reasoning from abstract assumptions to concrete empirical predictions can be extremely difficult, particularly about situations in which strategic interactions, expectations, beliefs, and communication work in subtle ways. The mathematics of comparative statics makes it possible for an analyst to check rigorously for mistakes in his or her own reasoning. It also makes it possible for others to confirm that the analyst's reasoning is correct—the theoretical equivalent of replication in empirical studies.

3. Testable hypotheses formally derived from explicit assumptions have the advantage of *unity.* In nonformal empirical work, analysts frequently sketch the reasoning behind multiple empirical conjectures. But the reasoning behind one conjecture may have little in common with that behind another. The two may even employ contradictory assumptions. Using a formal model to explicitly derive empirical hypotheses prevents this absurd situation.[8] It may also become clear which hypotheses can only be

7. Morton (1999) addresses the issues in this section more carefully, supplying many concrete illustrations.

8. An illustration: Lax and Cameron (2001) present a formal model of opinion assignment on the Supreme Court and formally derive a series of hypotheses. Many of these hypotheses already exist in the nonformal empirical literature but are motivated by a series of ad hoc and sometimes contradictory assumptions. In contrast, Lax and Cameron show that all can be derived from a single set of assumptions (assigners are effort-constrained maximizers of policy) and the same causal mechanisms.

derived from a different set of assumptions, opening the way to a test of competing models.[9]

Paradoxically, many political scientists seem to see the clarity, rigor, and unity of formal modeling as its principal *disadvantage* for empirical work. For some researchers, a fuzzy, possibly erroneous prediction from a nonformal model with unknown assumptions is somehow preferable. The unspoken belief seems to be that if you keep your assumptions implicit or hidden (perhaps even from yourself), you haven't made any assumptions! Of course, this position is utterly mistaken. Behind the selection of facts and narrative strategy of every case study and behind every regression in every quantitative analysis, lie a multitude of assumptions. Refusing to face those assumptions and their logical consequences is no virtue.

Toward Better Theory

In our exposition of SEM-like methods using the parallelogram model, we alluded several times to its inadequacy as theory. What's the problem, and what are the implications for FE work of building better theory?

The parallelogram model shoves most of the interesting politics in policymaking into the support and opposition functions, which are severely undertheorized. We concur with theorist David Austen-Smith's evaluation of the result:

> Questions about how and when influence is effective in majoritarian legislatures, about why some groups have "intrinsically more influence" than others, about how and why resources devoted by groups should map deterministically into a legislative decision and subsequent bureaucratic execution, about why groups adopt different patterns of activity (campaign contributions, informational lobbying, grass-roots activism, etc.), and so on, simply cannot be posed with the aggregate framework [i.e., an approach like the parallelogram model]. (1997, 299)

Austen-Smith concludes that better theory must be "micro-oriented" and more firmly grounded institutionally. Absent this kind of theory, it is hardly surprising that empirical estimates of parameters are unstable.

Austen-Smith's conception of better theory is widely shared among formal political theorists. It is exactly the direction contemporary formal theory is developing in American politics, comparative politics, and international relations. Within political economy, some researchers have preferred to retain very austere depictions of politics but embed them within quite elaborate models of economies, for example, of international trade. But others are moving to build models with strong micro- and institutional foundations.

9. Morton (1999) discusses alternative model evaluation at some length.

Models of this kind raise new issues for FE work. They are inevitably game theoretic. Comparative statics becomes subtler and more difficult.[10] The smooth, linear, continuous functions assumed in our expository example often go out the window. For example, in models using spatial theory, there can be distinct regimes. The behavior of political agents may differ dramatically across these regimes, but the conditions separating them may be difficult to observe empirically.[11] Problems like these raise a host of new and difficult methodological issues that need to be addressed in the years ahead.

Forging Stronger Links between Theory and Empirical Work

Our review of the current state of FE work in political science indicates that the average study consists of a simple formal model, used to gain insight into a phenomenon, followed by some empirical work (a case study or quantitative data analysis) loosely suggested by the model. The empirical work is inspired by the theory because the researcher uses the formal model to suggest relationships to look for and variables to employ. But the logical connection between theory and empirics is rarely closer than that. For example, the analyst may fail to derive formal comparative statics, consider the model's structural form, or ponder restrictions imposed on reduced forms. This inspired-by methodology gains the analyst the power of formal models for theorizing, which is no small matter. But it throws away the virtues of formal models for structuring empirical work. We predict FE work in political science will increasingly move away from the inspired-by methodology toward the SEM-like methods exposited above.

At the cutting edge of FE work, however, analysts are challenging certain elements of the SEM approach. We cannot hope to resolve the issues being raised, especially since many remain unsettled. But we can give the reader at least some sense of the debates.

Broadly speaking, two issues are at play:

- Is the formal model to be viewed as a *complete* model of the data generating process or as a *partial* model of the data in hand? The answer to this question determines the legitimacy of control variables and arbitrary assumptions about functional forms.

- What is the true nature of the stochastic element in the estimating equations, and what is its relationship to the formal model? The an-

10. A selection of essays addressing this point includes Dixit 1986, Fudenberg and Tirole 1984, and Hirshleifer and Rasmusen 1992.

11. A selection of essays addressing or illustrating this point includes Cameron, Segal, and Songer 2000, Lee and Porter 1984, Moraski and Shipan 1999, Segal and Wang 2001, and Spiller and Gely 1992.

swer to this question determines how the empirical evaluation of the formal model is to be interpreted.

Viewing a formal model as, at best, a partial picture of reality seems like common sense. Thus, it may seem natural to include control variables in the empirical analysis (control in the sense of multiple regression, not experimental manipulations), variables missing from the formal model. It may seem equally straightforward to specify tractable functional forms for estimating equations, so one can use standard statistical packages or results, even though one is unsure whether the specified form is compatible with the model's assumptions. But each such move drives a wedge between the formal model and the empirical analysis. At some point, the wedge becomes so large that an analyst is no longer using the formal model and certainly isn't testing it.

These problems have led some researchers to treat formal models *as if* they specify the complete data-generating process. In this view, if a variable isn't important enough to include in the formal model, it shouldn't be in the empirical work. Similarly, estimating equations should be strictly derived from the formal model, whose assumptions must be explicit enough and tractable enough to allow such derivations.

Similar issues arise concerning the stochastic elements in formal models versus estimating equations. Adding white noise error terms to deterministic models seems at best arbitrary and may lead to logical absurdities. For example, if the actors in a political situation understand that their world has a degree of randomness, this understanding is apt to affect their behavior. But if so, a deterministic formal model of the situation is simply wrong, and sprinkling white noise error terms in estimating equations won't fix it. This line of thought leads to incorporating stochastic elements directly in formal models (usually as games of imperfect or incomplete information) and carrying those stochastic elements through into the empirical estimation.[12]

These efforts may seem like an extraordinary effort simply to achieve logical consistency. Whether they are truly necessary remains controversial within the FE community.[13] In addition, such efforts often require considerable technical prowess. But they promise a payoff that political scientists have been slow to grasp. If full-blown, ultrarigorous structural models actually uncover stable causal parameters, *it becomes possible to perform theory-driven, data-sensitive policy experiments.*

12. Signorino (2000) provides a helpful analysis of these and related issues. In addition, Signorino has written software that estimates structural parameters for several commonly encountered strategic models, incorporating explicit, theoretically plausible stochastic elements. This software is currently available at http://www.rochester.edu/College/PSC/signorino/.

13. They have become quite common in economics, however. Illustrative examples are Donald and Paarsch 1996 and Rust 1994.

To see the point, refer again to figure 1. Let x_1 be a novel policy intervention (e.g., a change in a political institution, like voter registration requirements) with uncertain effect on y_1. If the *remaining* parameters in the model are truly stable, then one can use them to predict the effect of the policy intervention on outcome z, under different assumptions about the effect of x_1 on y_1. Theory-driven, data-grounded "what if" experiments about political and institutional reforms might well deserve the attention of citizens and policymakers. Where data and theory allow such "what if" experiments, the possibilities are exciting.

Conclusion

Nonformal empirical work can be informative and revealing. However, the move from nonformal empirical work to inspired-by studies lets the analyst tap into the power of formal models for the purposes of theorizing. The move from inspired-by work to SEM-like work gains the analyst clarity, rigor, and unity in empirical hypotheses. The move from SEM-like work to rigorous structural estimation opens the door to theory-driven, data-grounded policy and political analysis. In our view, each of these moves is valuable.

■ | Illustrative Examples

Space prohibits extensive consideration of real examples, which are invariably far more complex than the previous section's simplistic expository device. Instead, we supply some thumbnail sketches of illustrative examples that interested readers might wish to pursue. We group these into three categories. First are works that take seriously Austen-Smith's call for micro-oriented, institutionally rich theory seriously, and then match these models with SEM-like empirical methods. Second are works that retain somewhat summary models of politics but move toward rigorous structural estimation. Third are works that meld micro-oriented, institutionally rich theory with rigorous structural estimation. The selection of these examples is necessarily arbitrary; many more examples in all substantive fields of political science are compiled on the accompanying web page.

Micro/Institutionalist Theory, SEM-like Methods

American politics example

Filer, Kenny, and Morton (1993) examine the effect of income on voter turnout. They propose a game theoretic model in which highly motivated elites within social networks in turn motivate others to vote. The authors solve the elites' strategic turnout game and then formally derive compara-

tive static predictions about group turnout by income class. They test the formally derived predictions against country-level data from presidential elections.

COMPARATIVE POLITICS EXAMPLE

Huber and Shipan (2001) examine the efforts of legislators to control bureaucrats through the design of statutes, especially their specificity. The authors construct a formal model of statute design, incorporating substantial variation in institutional arrangements. They formally derive a series of comparative statics, focusing on issues like the extent of policy conflict between legislators and bureaucrats, the internal capacity of the legislature, conflict across chambers in bicameral legislatures, and the availability of nonstatutory means for controlling bureaucrats (e.g., legislative vetoes). They test the formally derived hypotheses with remarkable original data on the specificity of statutes. An interesting element of this analysis is its use of comparative data from the U.S. states simultaneously with cross-national data.

INTERNATIONAL RELATIONS EXAMPLE

Schultz (2001) uses FE work to examine information, threats, and war fighting—issues related to the Democratic Peace. The author builds a formal model of a unitary democratic state engaged in crisis bargaining. In the model, the monopoly party's actions convey information to domestic voters and a foreign bargaining partner. The critical comparative static comes from adding a domestic opposition party, whose support or opposition to the governing party also transmits information. Schultz shows that this institutional difference decreases the probability of war, decreases the likelihood that the democratic state makes threatening moves, and increases the effectiveness of the threatening moves it makes. Schultz then tests these and other propositions empirically, focusing on threatening moves and their effectiveness.[14]

INTERNATIONAL POLITICAL ECONOMY EXAMPLE

Mansfield, Milner, and Rosendorff (2000) examine the effect of regime type (democracy, autocracy) on the likelihood of concluding trade agreements. They construct a formal game theoretic model in which a government undertakes international agreements, which strategically convey information to voters about the government. The authors solve the game between the government and voters, allowing variation in the electoral

14. This choice is of methodological interest. The direct empirical evidence on the Democratic Peace appears inherently too weak to make a definitive determination of the existence and origins of this phenomenon (see, e.g., Gartzke 1998). Schultz's solution is to build a formal model that can accommodate the Democratic Peace but can itself be tested with different, less ambiguous data—an excellent demonstration of the power of joining formal models and empirical evaluation.

power of voters (hence, regime type.) They formally derive empirically testable hypotheses using comparative statics, which they test with data from recent trade agreements.

Austere Political Theory, Structural Estimation

INTERNATIONAL POLITICAL ECONOMY EXAMPLE

Goldberg and Maggi (1999) test one of the best-known models of endogenous tariff formation, the Grossman and Helpman model (1994), a lineal descendant of the early Chicago models. In this model, two interest groups simultaneously offer a government actor a schedule of campaign contributions, in exchange for trade protection. The government then chooses a level of tariffs. This somewhat austere model of politics is embedded in a fairly elaborate model of an economy. Goldberg and Maggi undertake a structural estimation of the Grossman and Helpman model. By solving the game between the groups and the government, they derive a specific functional form for the relationship between protection and key variables in the model. The stochastic element in the model is explicitly rationalized as measurement error. The authors estimate this equation using cross-sectional data from the United States for 1983, collected at the level of 3 digit SIC codes. The extensive use of sensitivity tests is a particularly interesting feature of the analysis.

Micro-Institutionalist Theory, Structural Estimation

AMERICAN POLITICS EXAMPLE

Schachar and Nalebuff (1999) use a structural approach to empirically evaluate a group-voting model with some similarities to Filer, Kenney, and Morton's model. Shachar and Nalebuff use their formal model to construct a maximum likelihood function for estimating turnout levels by state in presidential elections. The maximum likelihood function's form is directly derived from the formal model's equations. It is quite complex, incorporating a multitude of parameters in the formal model. An interesting feature of the analysis is the use of a structural model to conduct a modest "what if" experiment, concerning higher levels of turnout. They argue that if turnout levels had been 100 percent, Republicans would never have gained the presidency in the post–World War II era.

COMPARATIVE POLITICS EXAMPLE

Diermeier and Merlo (1999) consider a formal, stochastic bargaining model of government formation and duration in parliamentary democracies. The authors derive a parametric structural estimating equation that is compatible with the formal model. They estimate this with maximum likelihood. The data come from 236 governments in 9 countries over the pe-

riod 1947 to 1997. Using the estimated structural causal parameters, the authors undertake a series of policy experiments that examine the effects of different institutional rules (the investiture rule and the constructive vote of no confidence) on negotiation duration, government duration, and government size.

INTERNATIONAL RELATIONS EXAMPLE

Signorino and Tarar (2001) examine a straightforward incomplete information model of extended deterrence (that is, whether a country can prevent others from making war on its allies). They derive an estimating equation compatible with the formal strategic model, including the assumed form of incomplete information. They then structurally estimate the model using standard data sets on war. An interesting element of the paper is the use of the estimated structural causal parameters to interpret specific cases in recent history, for example, the Berlin blockade of 1948.

■ | Substantive Contributions of Formal Empirical Work

What are the substantive, as opposed to the methodological, contributions of FE work to political science?—for such is the basis on which FE work must ultimately be judged. We can only offer illustrations, for FE work has become so widespread and diverse that it extends far beyond our ability to evaluate knowledgeably. Nonetheless, these illustrations suggest that substantive accomplishments are relatively abundant and likely to grow.

In evaluating FE work, we use four standards:

1. *Understanding political phenomena and solving empirical puzzles.* Did the empirical work show that a particular formal model affords powerful analytic leverage over an important political phenomenon? In other words, did the FE work show the formal model has empirical punch?
2. *Advancing rich theory and stimulating new theory.* Did the empirical work lend support to a formal model or class of models or lead to the formulation or refinement of new theories or models? In other words, does the FE work lend credence or stimulate formal models with theoretical punch?
3. *Rejecting theory.* Conversely, did the empirical work allow the rejection of a plausible formal model or class of models, as offering relatively little empirical or theoretical leverage over an important political phenomenon?
4. *Improving public policy.* Did the FE work offer convincing grounds for better public policy? Did it work as applied political science?

SOLVING PUZZLES AND UNDERSTANDING THE WORLD OF POLITICS

- *The activity puzzle in congressional studies.* In legislative studies, a puzzle of the 1980s was statistical evidence indicating that congressmen's activities (e.g., constituency service) did not affect their reelection prospects. The FE work in Rivers and Fiorina 1989, employing structural estimation, solved the puzzle.
- *Outlier committees in Congress.* Determining the extent and distribution of outlier committees (ones in which the ideological or policy preferences of members differ significantly from those of the average member) was a central thrust of congressional scholarship in the 1990s. This work was stimulated by Gilligan and Krehbiel's formal models (1990) of legislative organization and Krehbiel's subsequent FE work (1990). The work led to a better understanding of the phenomenon.
- *Duverger's law.* Gary Cox's FE work on party structures remains a touchstone in comparative politics.
- *Veto politics.* Recent FE work on vetoes and veto threats arguably advanced the understanding of the presidential veto and interbranch bargaining in separation of powers systems (C. M. Cameron 2000; Groseclose and McCarty 2001).
- *Gridlock in separation of powers systems.* Krehbiel's FE work (1998) on policymaking in the United States establishes the standard for future work in this area.
- *Legislative delegation to agencies.* Epstein and O'Halloran's FE work (1999) established a coherent theoretical framework, demonstrated the feasibility of systematic empirical study, and documented plausible patterns in legislative delegation to agencies.
- *Referenda and legislative constraints.* E. R. Gerber's FE work (1999) on state referenda shows that the presence of this device forces legislators to remain more proximal to the desires of the median voter in the state.

ADVANCING RICH THEORY

- *The spatial model of elections.* The intellectual elegance of the spatial theory is clear. But FE work helped establish its practical usefulness and in turn further stimulated theorists (Enelow and Hinich 1984). But see "Rejecting Theory," below.
- *Duvergerian conceptions of party structure.* Again, primarily via Cox's landmark work. These theories have become a staple and an intellectual highlight of comparative politics.

- *The monopoly agenda setter model.* Work in FE employing and extending Romer and Rosenthal's monopoly agenda setter model confirms that it is a fundamental tool for thinking about separation of powers systems.
- *Regulatory politics.* The FE work of Chicago political economists fundamentally altered discussions of the subject and exerts considerable influence in IPE studies of tariffs and trade barriers.
- *Informational foundations of war.* Drawing on simple formal theory, Fearon (1995b) argued that war between states is most likely to result from informational, rather than material, factors. This mode of thinking represented a considerable departure from realist views (including ones in early formal models). Ongoing FE work has considerably elaborated and empirically tested these ideas.

Rejecting Theory

- *Downsian models of party politics.* Work in FE has consistently found a substantial degree of policy divergence between candidates, platforms, or policy proposals, even in circumstances that seem to approximate the conditions for the median voter theorem.
- *The separation-of-powers model of judicial independence.* The case is hardly closed, but accumulating evidence suggests that federal courts in the United States are not as responsive to Congress as simple separation-of-powers models seem to suggest (Segal 1997).
- *Political-business-cycle models.* This is a very complex area, but evidence seems to reject early, naïve theories, though the jury is still out on more-sophisticated models.
- *Commitment model of veto threats.* The evidence on veto threats clearly rejects the simplest version of commitment-type veto threats as a general explanation of how veto threats work.

Improving Public Policy

We know of few areas in which contemporary academic political science of any kind has demonstrably affected public policy. Consequently, we simply note a few FE research programs that have the clear *potential* to improve public policy.

- *Juries.* Formal work has suggested that changing jury voting rules might lead to different and perhaps better decision making. Experimental evaluation of different rules is under way (Guarnaschelli et al. 2000). If any of the modifications were adopted, the impact on millions of jurors as well as litigants might be enormous.

- *Campaign finance.* Ongoing FE work (e.g., McCarty and Rothenberg 1996) has obvious policy implications.
- *Sequential elections versus simultaneous elections.* Investigations in FE of sequential elections (e.g., as in presidential primaries sequenced over many states) versus simultaneous elections (e.g., one large national primary) have obvious policy implications (Morton and Williams 2001).
- *Redistricting.* Recent FE work on the policy consequences of racial redistricting has substantial policy implications (Shotts 2001).
- *Extended deterrence in international relations.* Signorino's use (1999) of a structural model to interpret historical cases of extended deterrence raises the possibility of using this and similar structural models for predictive and policy analytic purposes.

Summary: The Trajectory of FE Work

We believe the trajectory of FE work is illustrated by advances in an area one of us knows well, turnout in elections. Turnout is a notoriously difficult topic for formal theory; hence, the growth of FE work in such stony soil provides a kind of critical test.

In this area, theorists first formulated a simple, deterministic model of turnout.[15] This model made stark point predictions, allowing evaluation through simple observation of errant observations. These were notoriously abundant. Yet, the initial, brutal collision between theory and data did not lead theorists to abandon their general framework. Instead, most formal theorists temporarily bracketed their failure to explain turnout and used the same general framework to study electoral choices, where it worked much better. Meanwhile, other theorists developed strategic models of the calculus of voting, incorporating the role of groups and the probabilistic nature of leader influence (see the discussion of Filer, Kenney, and Morton 1993, above). Empirical estimations in the SEM style supported the models' predictions. However, the initial rounds of evaluation of the second-generation models stopped short of rigorous structural estimation. Finally, structural estimation of a formal model—using empirical equations directly derived from the formal model's equations—showed support for the formal model as an explanation of voter turnout decisions, explaining nearly 90 percent of the variation in turnout (see Schachar and Nalebuff 1999, above). The structurally estimated models allowed analysts to make theoretically driven, empirically grounded political and policy predictions. Needless to say, this work is not the last word on the subject, but the trajectory in this difficult area is impressive.

15. For a review of the early turnout literature, see Morton 1991.

■ | Conclusion

We are both assertive and modest about the value of FE work. We are assertive in claiming that a thorough, ongoing confrontation between formal theory and data will improve both theory and empiricism in political science and take the discipline a step closer to a cumulative, sophisticated social science of politics. It will improve formal theory by providing feedback on which models have empirical bite and which do not, and thereby stimulate new theorizing along productive lines. It will improve empirical work by allowing analysts to exploit the clarity, rigor, and unity of formal models as devices for generating hypotheses and help them escape the degenerating inductivism so well parodied by the cognitive scientist Allen Newell in his essay, "You Can't Play 20 Questions with Mother Nature and Win." Finally, by providing the profession with a core of clear, deductive models of proven empirical bite, it will advance the discipline as a whole and gradually open the door to theoretically driven, empirically grounded political analysis.

We are assertive in calling for more (and better) FE work in empirical political science, but we also wish to be modest, for two reasons. First—simply to be clear—we do not claim, or for a moment believe, that FE work is the only way to go. The well-established, older style of descriptive and inductive empirical work—quantitative and qualitative—will continue to be essential for providing knowledge about the world of politics. In our view, the only sensible position is that nonformal and formal empirical analyses both contribute to a more robust political science. Similarly, pure formal theory—theory unaccompanied by data and statistical tests—is a precious commodity. Investing in formal theory is investing in our discipline's basic intellectual infrastructure. In fact, formal theory assumes an even greater importance when political scientists are serious about confronting theory with data. But a second and equally important cause for modesty is the difficulty of the enterprise. Combining theory and data involves more than mastering two different skill sets. It requires new ways of thinking and solutions to new methodological problems, problems that political scientists have hardly begun to face. Solving these problems will be a major challenge for the new century of political science.