PART II

THEORY AND METHODOLOGY

CHAPTER 2

CONTEXT MATTERS: THE CHALLENGE OF MULTICAUSALITY, CONTEXT-CONDITIONALITY, AND ENDOGENEITY FOR EMPIRICAL EVALUATION OF POSITIVE THEORY IN COMPARATIVE POLITICS

ROBERT J. FRANZESE, JR.

1. Introduction: The Centrality of Multicausality, Context-Conditionality, and Endogeneity to Classical and Modern Comparative Politics

Modern scholarly reviewers characterize[1] the pre-war and immediate post-war study of comparative politics as legalistic, favoring categorical enumeration over positive-theoretical analysis of constitutional details, and as parochial and, indeed, as non-comparative, exhibiting Western (often specifically US) bias in the topics studied and in normative conclusions and rather lacking in theoretical or empirical comparison.

From the mid-1950s, Gabriel Almond (1956) and contemporaries, applying a Parsonian approach to social science, led a political sociology revolution in comparative politics. Sparked by the catastrophic rise of fascism and dictatorship that plunged the globe into war, and by democracy's failure to advance and secure its initial post-war successes, the central question for these scholars was what conditions fostered stable, democratic political development. Inspired by contemporary scientific sociology, they sought answers in the polity's social structure: e.g., its homogeneity and stratification (Almond 1956), its socioeconomic development, or the cross-cutting or reinforcing nature of its sociopolitical cleavages (Lipset 1960). Perhaps most notable about this revolution was the movement it signaled from configurative description toward a positive science of comparative politics that asks theoretical research questions (e.g., what societal characteristics may contribute to democratic development and stability, and how?) and not merely descriptive (e.g., what does the French constitution say?) or historical-factual (e.g., who voted for Hitler?) ones and that proposes positive theories about causal relationships in answer rather than unadulterated parochialism or bias or normative judgement. Empirical evaluation of these positive theories, however, remained depressingly impressionistic and, perhaps, too often as parochial and biased as earlier configurative descriptions had been.

The political culture and political behavior revolutions of the 1960–1970s completed the movement in comparative politics from configurative description to positive social science. Almond and Verba's (1963) classic *Civic Culture* perhaps initiated and still best exemplifies both revolutions in following the posing of *a positive question*—what fosters stable, well-functioning democracy (which they defined precisely enough)—with *logically argued, positive-theoretical, hypothetical answers*—crudely: a citizenry with beneficial cognitive, affective, and evaluative

[1] The ensuing intellectual historiography of comparative politics as a field of enquiry is likely more caricature than characterization. It serves here merely to provide background for how the notion that *context matters* is and always has been a core tenet of the field.

orientation toward the political system (which they defined precisely enough)—*and empirical evaluation* based on equally sufficiently precise and objective measurement of key components (variables) in the argument. *The Civic Culture*, however, suffered a critical limitation that its ultimate explanandum (dependent variable), stable, well-functioning democracy, remained impressionistically measured and in only five contexts (nations). For this reason, the book more solidly established the extent and content of the *Civic Culture* in these five nations than it did the posited theoretical (causal) relationship between *Civic Culture* and well-functioning democracy. Later work in the cultural-behavioral tradition, e.g., Inglehart's (1990) *Culture Shift*, reduced these limitations, in the process perhaps cementing the case, begun by pioneers like Karl Deutsch (1971), for the utility of large-sample statistical analysis to the positive study of comparative politics.

By the 1980s, social structure, political culture, and public opinion and behavior had become the main sources of likely independent variables in the modern, positive, political science of comparative politics, and statistical analysis of comparative-historical data had become one important tool in empirical evaluation of those positive arguments. However, this tool also enabled scholars to discern that, in fact, social structure seemed to determine political outcome—e.g., social homo- and heterogeneity related to (in)stability (Powell 1982), societal fractionalization and polarization related to party system (Sartori 1976), etc.—less fully, universally, and surely than previously thought. Spurred by the weakness or incompleteness of such social structural explanations for macro-political outcomes and perhaps also unsatisfied with the immediacy and causal proximity with which culture, beliefs, attitudes, and opinion linked to each other and to micro-behaviors like, say, vote choices, Sartori, Powell, Smith (1972), Berger (1981), Lehmbruch and Schmitter (1982), Lijphart (1984), and others returned institutions—political, social, and economic—to the center of analysis. Building from earlier work that theoretically and empirically linked, e.g., electoral law to party system outcomes (Rae 1967) and party and governmental systems to coalition politics (Riker 1962; Dodd 1976), these authors argued socioeconomic structure *works through* political, social, and economic institutions to shape the incentives of political actors: voters, workers and employers, policy-making and party elites. Comparative-historical statistical analysis again helped establish these claims empirically, showing that, *in addition to* or *controlling for* socioeconomic-cultural conditions, presidential, majoritarian-parliamentary, and representative-parliamentary institutions affect participation and social and governmental stability (Powell 1982), institutional structures of labor help determine political-economic performance (Cameron 1984), majoritarian or consensual institutions shape democracies' performance (Lijphart 1984), etc. This effectively added sociopoliticoeconomic institutions to the growing list of (classes of) key explanatory variables. However, the full recognition of the implications of socio-economic-cultural conditions *working through* institutions, which implies that the effects of institutions *depend on* these conditions and, vice versa, that the effects of socioeconomic-cultural conditions in turn depend on those institutions, went largely unexplored in statistical empirical work for some time.

The modern, positive-theoretical study of comparative politics thus emphasizes the societal structure of interests, political culture and public opinion, and sociopoliticoeconomic institutions, in explaining the intranational, cross-national, international, and/or cross-temporal variation observed in political outcomes. In this regard, the field has come full circle. The central tenet of modern comparative politics is, as that of classical pre- and post-war comparative politics was, that *context*—structural, cultural, institutional, and strategic; social, economic, and political; international, domestic, and local—matters. More precisely, *context matters* in at least three ways. First, the outcomes we seek to explain, understand, or predict have multiple causes, so the values of the many potential causes in any given context affect the outcomes: *multicausality*. Second, the effects of each cause on outcomes tend to vary across contexts, which is to say that the effects of each cause tend to depend on the values of one or more other potential cause(s) present in that context: *context-conditionality*.[2] Third, the many outcomes and many putative causes in the political world that we seek to understand tend, in fact, to cause each other to some degree rather than some factors being only causes and others being only effects: *endogeneity* (synonyms: simultaneity, reverse causality, bi- and multidirectional causality).

These three aspects of the "*context matters*" central tenet of comparative politics—*multicausality*, *context-conditionality*, and *endogeneity*—also represent three of the most ubiquitous and severe challenges to empirical inference in political science. Indeed, although perhaps most directly implied by the *context matters* mantra of comparative politics, *multicausality*, *context-conditionality*, and *endogeneity*, along with the relative paucity of information—that is, too few observations: we typically have, after all, only the one comparative history of the world from which to infer anything—are perhaps the central challenges to empirical evaluation across all of social science (and in many if not all natural sciences as well).[3]

In short: first, almost everything matters (i.e., many X's cause most of the Y's studied throughout social science); second, how each X matters depends on almost everything else (i.e., the effects of each X on some Y typically depends on many other X in that context); and, third, everything pretty much causes everything else in sociopoliticoeconomic reality (i.e., almost everything in society, polity, and economy is endogenous to almost everything else in and across those spheres). Finally, to make matters worse, we usually have precious little empirical information with which to sort through all this complexity.

That *context matters* in these ways is sometimes taken as a challenge for statistical methods of empirical evaluation of theory in particular, but the challenges are logically inherent to the substantive propositions of multicausality, context-conditionality, and endogeneity and do not inhere, therefore, to the particular empirical-methodological approach taken to (partially) redress them. Stated differently, these challenges do not arise because specifying a statistical model, i.e., writing one's

[2] This includes history, and so context-conditionality subsumes historical path and state dependence.

[3] *Comparative Politics*, as a colleague is fond of (correctly) saying, is a subject matter and not a methodology (W. Clark: personal communications); that a chapter on methodology in a handbook of comparative politics should address methodological concerns of at least social scientific breadth is therefore wholly fitting.

empirical arguments formulaically, highlights and clarifies them mathematically, and they do not go away simply because one neglects to do so. Likewise, the challenges do not arise because some scholar records the information she observes numerically and analyzes them statistically as observations in a dataset, and they do not disappear if some other scholar instead records the information he observes qualitatively and analyzes it in some manner as "causal-process observations" (Brady and Collier 2004; *Political Analysis* 2006: 14 (3)). Furthermore, as shown below, the challenges are not necessarily surmounted, nor indeed are they often surmountable even in principle, solely by analyzing some available empirical information more closely or simply by gathering more empirical information. This is because the challenges are logical and theoretical as much as, or more than, empirical. Thus, these are the challenges of empirical evaluation in social science, and not those of quantitative or qualitative methods, and, insofar as we manage to learn something from our empirical analyses,[4] quantitative or qualitative, we must somehow have redressed these challenges to some degree.

That quantitative and qualitative empirical studies face the same logical challenges is now widely accepted, and some very useful works (e.g., King, Keohane, and Verba 1994; Brady and Collier 2004) have begun to consider how the approaches may be understood from this perspective and how analyses of each sort may be improved by understandings gained from the other. Relatively missing from these useful discussions, contributions, and debates, however, has been explicit statement from the statistical perspective of these fundamental challenges[5] that both approaches face and discussion of the choices each must make as necessary conditions to learn from comparative history. The rest of this chapter offers such explicit discussion so that the formal statement of the challenges from this perspective will help researchers from both perspectives understand more fully the challenges they face and the choices and tradeoffs they make in redressing them.

2. The Problem of Too Few Observations/Too Little Information: Quality vs. Quantity

Before proceeding to consider multicausality, context-conditionality, and endogeneity, an abstract consideration of the terms of the tradeoff between the quality and the quantity of information brought to bear upon a question of empirical inference may

[4] "Learn something" and similar such phrases below mean "learn something helpful in general empirical evaluation of positive theory." One can of course learn many useful things on many other dimensions from empirical description that is useless for general empirical evaluation of positive theory.

[5] Certainly, formal statements of statistical models or discussion of their use and implementation are not in short supply. What is missing has been an explicit formal statement from the statistical perspective of the challenges for empirical analysis of comparative politics and discussion of the terms of tradeoffs researchers in that substantive area must make.

be enlightening. Given the constraints of time, competencies, and the availability of information, researchers often must choose between observing more pieces of information more cursorily and fewer pieces of information more fully and accurately.[6] The terms of this tradeoff are impossible to determine with great precision as a general proposition, but we can offer some help to gauge those terms broadly by considering the tradeoff between the accuracy of some measures and the number of such measures used to estimate some quantity of interest.

Suppose, for example, that some researcher is interested in the empirical relation between the quality of democracy and the level of economic development in some society. Suppose further that the actual relationship between economic development, *EcDev*, and the true quality of democracy (in whatever meaningful sense), *QualDem**, is the following:

$$QualDem^* = \beta \times EcDev + \epsilon \tag{1}$$

where ϵ is some random noise, with variance σ^2_ϵ, since the relationship is not deterministic and exact.[7] Now suppose, realistically, that the researcher can measure the true quality of democracy, *QaulDem**, only with some error, γ, whose variance, σ^2_γ, he can reduce by focusing his empirical studies on fewer contexts. This means that the researcher can only evaluate empirically the following relationship:

$$QualDem = QualDem^* + \gamma \Rightarrow QualDem = \beta \times EcDev + \epsilon - \gamma \tag{2}$$

Under the usual conditions, the researcher obtains the best[8] estimate of β by:

$$\hat{\beta} = \frac{Cov(QualDem,\ EcDev)}{Var(EcDev)} \tag{3}$$

and this best estimate of the relationship between economic development and the quality of democracy will have a variance (i.e., uncertainty) of

$$V(\hat{\beta}) = \frac{\sigma^2_\epsilon + \sigma^2_\gamma}{Var(EcDev)} = \frac{\sigma^2_\epsilon + \sigma^2_\gamma}{(n-1) \times \sigma^2_x} \tag{4}$$

Thus, to characterize the terms of the tradeoff between the quantity of information and the accuracy of the measurement of that information requires that one be able to gauge the relative contributions to uncertainty regarding the relationship of (a) measurement error, σ^2_γ, (b) the inherent uncertainty in the actual relationship, σ^2_ϵ,

[6] The phrase *pieces of information* intentionally replaces the more common *observations*, *cases*, or *countries*, because the issues raised and discussed do not hinge in any way on the comparative-historical analysis under consideration being within or across countries or cases and because whether the information gleaned is labeled a dataset or a causal process observation is likewise irrelevant to the present discussion.

[7] We use an extremely simple bivariate and linear model with additively separable error component here strictly for expositional convenience. The results of this consideration of weighing quality vs. quantity would be complicated but not changed in upshot if these simplifications were abandoned.

[8] *Best* here means BUE, the best unbiased estimator, under the usual conditions (and the restriction to the class of linear estimators being unnecessary because we have stipulated that the true relationship is indeed linear).

and (c) the variation in the explanatory variable, σ_x^2, which last determines the (expected) contribution to the certainty of the estimate of each piece of information.[9] Researchers wondering how to trade quantity for quality will not know the values of these crucial quantities a priori, of course—indeed, absent further theory or assumption, distinct estimation of σ_γ^2 and σ_ϵ^2 is impossible—but a sense of the expected variation in the explanatory factor, σ_x^2, might be obtained from data (plus assumptions) in some cases. Researchers could, in any event, profit from considering the equation, inserting their own substantive-theoretical senses of the relative magnitudes of explanatory variable variation, inherent variance (i.e., uncertainty or randomness) in the outcome (given their model), and measurement error variance.

Table 2.1 gives some examples where inherent uncertainty in the relationship, σ_ϵ^2, is one, and measurement uncertainty varies across the columns from one-tenth of that,

Table 2.1 Example terms of quality vs. quantity tradeoffs

	Variation of explanatory variable (σ_x^2): 0.5						
	$\sigma_\gamma^2 = 0.1$	$\sigma_\gamma^2 = 0.25$	$\sigma_\gamma^2 = 0.5$	$\sigma_\gamma^2 = 1$	$\sigma_\gamma^2 = 2$	$\sigma_\gamma^2 = 4$	$\sigma_\gamma^2 = 10$
n=2	2.200	2.500	3.000	4.000	6.000	10.000	22.000
n=10	0.244	0.278	0.333	0.444	0.667	1.111	2.444
n=50	0.045	0.051	0.061	0.082	0.122	0.204	0.449
n=250	0.009	0.010	0.012	0.016	0.024	0.040	0.088
	Variation of explanatory variable (σ_x^2): 1						
	$\sigma_\gamma^2 = 0.1$	$\sigma_\gamma^2 = 0.25$	$\sigma_\gamma^2 = 0.5$	$\sigma_\gamma^2 = 1$	$\sigma_\gamma^2 = 2$	$\sigma_\gamma^2 = 4$	$\sigma_\gamma^2 = 10$
n=2	1.100	1.250	1.500	2.000	3.000	5.000	11.000
n=10	0.122	0.139	0.167	0.222	0.333	0.556	1.222
n=50	0.022	0.026	0.031	0.041	0.061	0.102	0.224
n=250	0.004	0.005	0.006	0.008	0.012	0.020	0.044
	Variation of explanatory variable (σ_x^2): 2						
	$\sigma_\gamma^2 = 0.1$	$\sigma_\gamma^2 = 0.25$	$\sigma_\gamma^2 = 0.5$	$\sigma_\gamma^2 = 1$	$\sigma_\gamma^2 = 2$	$\sigma_\gamma^2 = 4$	$\sigma_\gamma^2 = 10$
n=2	0.550	0.625	0.750	1.000	1.500	2.500	5.500
n=10	0.061	0.069	0.083	0.111	0.167	0.278	0.611
n=50	0.011	0.013	0.015	0.020	0.031	0.051	0.112
n=250	0.002	0.003	0.003	0.004	0.006	0.010	0.022

[9] The example simplifies matters dramatically by confining measurement error to the dependent variable, by considering only one explanatory factor, by omitting an intercept from the true relationship, and by finessing the difference between the variance of a random variable and the sample variation in a

so that lack of information quality is just less than 10 percent of the total numerator uncertainty, to ten times that, so lack of quality is just over 90 percent of that total. The top section of the table considers a case with relatively little variation to be gained from greater quantities of information, $\sigma_x^2 = .5$, so independent variable variation is only half the true variance of the dependent variable, to a case at the bottom with relatively much to gain, $\sigma_x^2 = 2$, so independent variable variation is twice the dependent variable's true variance, with a case in between where independent variable variation equals dependent variable variance. Down the rows, the table lists quintuplings of the sample size (i.e., amounts of information) from 2 to 10 to 50 to 250. The cell entries give the uncertainty (namely, the variance) of the estimated relationship between the independent and dependent variables. The table reveals, e.g., that, under the most extremely low-quality information conditions considered (the last column, $\sigma_\gamma^2 = 10$), to compensate for an 80 percent reduction in information from 50 to 10, one would need a 90 percent improvement in information quality (to $\sigma_\gamma^2 = 1$), which might seem reasonable. To compensate for a further 80 percent reduction in information from 10 to 2, however, would require a ninety fold enhancement of information quality. To restate it more strikingly from the other side, an increase in sample size from the lowest possible, 2,[10] to a still relatively small 10 would increase the quality of the inferences the researcher could draw as long as the quality of information worsened by no more than 9,000 percent (i.e., 90×) or so, from a very high-quality $\sigma_\gamma^2 = .1$ to a situation where measurement error was about 90 percent of total uncertainty about the dependent variable ($\sigma_\gamma^2 \approx 9$). Judging from the table, further increases in sample size, from 10 to 50 and from 50 to 250, likewise merit undertaking as long as information quality does not diminish by too much more than 4,000 percent (40×) to something over $\sigma_\gamma^2 = 4$, i.e., greater than 80 percent of the total uncertainty of the researcher about the relationship (but not to nearly as close to $\sigma_\gamma^2 = 10$ and over 90 percent as in the first case). Furthermore, the comparisons would look dramatically worse again for quality if reductions in quantity entailed also a reduction in useful independent variable variation (movements up the table). In short, over the ranges of σ_γ^2, σ_ϵ^2, and σ_x^2 that comparativists likely typically travel, information quality in larger samples must be abysmal and represent overwhelming proportions of researchers' total uncertainty (i.e., many times the uncertainty related to shortcomings in theory plus the inherent randomness of outcomes), and reductions in the quantity of information must come at quite

regressor. The insights generated hopefully merit the simplifications, but, for the record: (1) including a constant would add nothing of interest to the discussion; (2) considering error in the explanatory factor(s) would add a bias cost to the inefficiency discussed in the text, but its magnitude would simply parallel that of the inefficiency, although (3) the magnitudes of the bias and efficiency costs of measurement error(s) where more than one explanatory factor enters would be far more complicated to determine (see Achen 1985, 2002); and (4) the finessing merely allows the discussion to sidestep explicit consideration of the expected contribution of an additional observation, which would require a long discussion before arriving to the $(n-1) \times \sigma_x^2$ denominator as given.

[10] Inference about the relationship between 2 variables from 1 piece of information on the pair is obviously impossible. Graphically, that corresponds to drawing a line through one point; infinite such lines exist.

benign costs in lost variety (i.e., useful variation) and quite high gains in quality, if trades of quantity for quality are to offer gains in researchers' certainty about the relationships they wish to explore empirically. Clearly, then, a case for narrowing one's focus to fewer contexts solely on the basis of enhancing the quality of the information from those fewer contexts is extremely difficult to sustain.[11]

Proponents of pursuing greater depth of observation from fewer contexts, therefore, must see advantages beyond simple quality-of-information improvements. Potentially valid candidates are not lacking; indeed, as we will conclude and explain later, *qualitative analysis is an essential and valuable part of the scientific enterprise*. The above discussion simply demonstrates that one cannot easily sustain an argument that researchers should confine their attention more deeply to fewer contexts on the *sole* basis of the greater quality of information that surely affords. Certainly, e.g., the greater breadth of information that researchers may obtain when looking more closely at fewer contexts can essentially multiply observations within contexts (King, Keohane, and Verba 1994), although we should now reiterate that these extra pieces of information from within one context may not offer as much useful variation in explanatory factors (σ_x^2) as would the same quantity of additional information from other contexts. (Traversing contexts may raise other challenges, though, as discussed later.) Many scholars, however, have not much appreciated King et al.'s view of qualitative research in fewer contexts, which stressed efforts and attention to multiplying observations within those contexts (and to case selection), as it might seem to have relegated qualitative empirical analysis to the role of a necessarily inferior substitute when statistical empirical analysis is truly not possible, a substitute whose inherent inferiority is to be minimized by the former better approximating the latter.[12] Accordingly, numerous other advantages, such as some usefully greater ability to trace processes, and possibly thereby to grapple with multicausality and/or context-conditionality or to identify causality better, have also been claimed (Hall 2003; Brady and Collier 2004; *Political Analysis* 2006, 14 (3)).

Deferring discussion of such arguments until we turn, as we will next, to consider multicausality, context-conditionality, and endogeneity in sequence, let us first consider how, in light of foregoing considerations, we might interpret claims sometimes made (e.g., Rogowski 2004) for the empirical power and utility of single-case studies over and above any intra-case multiplying of observations. First, of course, empirical *utility* encompasses more than precise estimation of relationships. Single-case studies clearly can serve to raise theoretical conjectures and hypotheses for further exploration, for example, as has long been recognized (Przeworski and Teune 1970). Notably in this regard, e.g., Rogowski is careful to express all the empirical claims derived from his considerations of prominent single-observation studies in the conjunctive tense. Moreover, the points he makes that these studies of

[11] This conclusion, moreover, does not depend on the simplicity or linearity of the model, nor to any great extent on the particulars of the tabulated examples (see n. 9).

[12] Noting that statistical analysis is always possible, just as qualitative analysis is, if the *logically* necessary conditions for revealing empirical analysis have been met, would surely only heighten these sensitivities.

anomalies challenge previously held views of the empirical support for particular theories rest on knowing where the single observations in question lie relative to the rest of an implicit sample, either decidedly off the pattern of independent and dependent variables set by those only implicitly referenced others or far from the centroid (i.e., the multidimensional mean) of the independent variables in those others but not with dependent variable taking values where simple (usually linear) extrapolation of the theory to those extremes would predict. Discovering an extreme case that does not remotely fit implicitly referenced patterns from prior studies or that lies far from where extrapolation of the theories under consideration would predict is just what Rogowski argues it is: *potential* empirical anomaly[13] that should motivate theoretical refinement. As such, it is indeed potentially empirically powerful and useful, but it is not (nor does or should it claim to be) general empirical evaluation of theory in the same sense as would be discovering statistical unlikelihood that the pattern across several contexts follows the theoretical predictions.[14]

3. Multicausality: Almost Everything Matters

We turn now to the first of the three fundamental challenges for empirical analysis raised by the mantra that context matters (after the "too little information" problem which pervades and exacerbates them all): *multicausality*, or the conviction that many possible causal factors potentially operate in any given context. In one (not so extreme) example, Huntington (1991) identifies at least twenty-seven explanatory factors in democratization accounts in the literature before naming five most compelling: (1) deepening legitimacy problems of authoritarian systems as democratic norms are becoming increasingly globally accepted; (2) economic growth and expanded middle classes; (3) transformation of churches (especially the Catholic

[13] So-called *critical cases* do not exist unless one believes the socioeconomic-political world deterministic, i.e., as long as one considers outcomes at least partially driven by chance. In general, in a stochastic world, any one observation, no matter how extreme, could always have arisen by chance (unless the chance is bounded more tightly than the apparent anomaly in the outcome's extremity). This also means that two observations would suffice for Mill's method of determining causality by comparing observations alike in every detail but one putative cause and (potentially) one putative effect only in a deterministic and not in a stochastic world.

[14] The more apt analogy in statistical methodology is to outlier detection and sensitivity analyses, which may serve to begin to justify the contention above that qualitative analysis is an essential part of the scientific enterprise. The process and product of these statistical analogs, however, also help explain how and why the reliance on a single case and only implicit reference to some background empirical information set leaves the researcher reliant on discovering *extreme* departures from predictions or from the mean elsewhere for any reasonable confidence that the perceived anomaly is indeed anomalous.

Church) from defenders of the status quo to opponents of authoritarianism; (4) changes in the policies of external actors (USA and Europe in particular); and (5) snowball effects (i.e., diffusion). Dahl (1971) lists eight (pre-)conditions for democracy: the peaceful evolution of democracy (yielding clean transfer of legitimacy from the old to new regime), decentralized economy (avoiding economic power concentration), economic development, economic equality, social homogeneity, elite pro-democratic beliefs (ideally with authority structures similarly democratic across societal institutions), popular beliefs in democratic efficacy and in the sincere intentions of adversaries, and passive or supportive international conditions. Even leaving for the next section the complex *context-conditionality* inherent in several of these contentions, democratization theories, taken together (plus controls)—and taking them separately (or omitting important controls) would be dangerous for omitted variable reasons discussed below—offer quite an array of putative causes and so quite a complexly contextual account of the causes of democratization.

Let us again follow a strategy of discussing from an econometric perspective a simple example illustrating the challenges such multicausality raises for empirical 'evaluation: those of controlling potentially confounding covarying factors and determining their relative (partial) explanatory impact, and those of omitted variable bias, "included-variable bias,"[15] and Achen's (2002) "rule of three." Let us assume for simplifying expositional purposes that the researcher determines that just two factors should favor democratization (*Dem*): the level of economic development, *EcDev*, and of equality in its distribution, *EqDist*. Furthermore, development and equality matter simply linear additively (e.g., they do not interact), no other factors enter, and the inherent randomness in democratization outcomes is additively separable from the systematic component that depends on *EcDev* and *EqDist*:

$$Dem = \beta_1 EcDev + \beta_2 EqDist + \epsilon \tag{5}$$

In such a case, a simple Venn diagram can usefully—if incompletely, imperfectly, and not fully generally (see, e.g., Ip 2001)—depict the challenge of discerning and distinctly gauging the effects of economic development and of equality. In Figure 2.1, e.g., the areas of the oval labeled X_1 could depict the total variation observed across cases in democratization outcomes, and the ovals X_2 and X_3 those of *EcDev* and *EqDist* respectively. The overlapping sections reflect covariation of the corresponding variables. So, for example, the coefficient relating X_2 to X_1, following the formula from (3) above is given by the area of [2]+[3], depicting covariation of X_2 and X_1, divided by the are of the oval depicting X_1's variation, [1]+[2]+[3]+[4]. In multicausal contexts, generally, this will not be a good description of the variation in X_1 "due to" X_2 because [3] reflects the extent to which the outcome associates also with X_3—called shared covariation (and in a different way because [4] is variation in X_1 "not available to be due to" X_2). If information sets that isolate just one single

[15] The term, explained below, is borrowed from lectures by Gary King the author was privileged to hear.

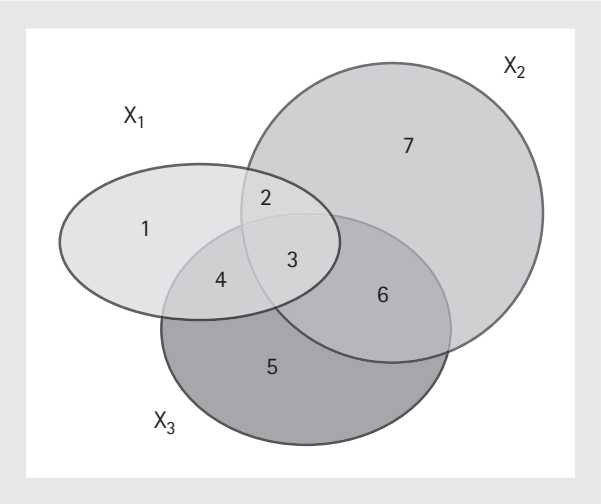

Fig. 2.1 Stylized Representation of Variation, Covariation, and Partial Coefficients in Multiple Regression

potential cause cannot be created—and they cannot as long as multicausality exists[16]—then some strategy of isolating the shares of covariations and variations due to each potential cause net of those due to others must be adopted. An important aspect of the challenge here is that no empirical information exists to inform allocation of [3], the covariation with X_1 shared by X_2 and X_3, across those covariates. One can theoretically impose an ordering or an allocation—as in path analysis or stepwise regression—and this may sometimes be useful and informative, but one cannot determine how to allocate [3] from empirics alone.

Multiple regression and related statistical analyses resolve this by basing estimates of the relationships only on the unique variation and covariation of the variables. That is, the partial association of X_2 with X_1 (i.e., the *partial coefficient* on X_2 in this model and, under the current assumptions, the *effect* of X_2 on X_1 *controlling for* X_3, i.e., after netting the relationship between X_3 and X_1) is [2]/[1] +[2]. This seemingly simple procedure is one crucially important contribution of multiple regression to comparative empirical analysis. Since, for the most part, we have only comparative history as our database, i.e., we have only the world as it is and was, to control for multiple possible causes by holding all potential causes exactly fixed except the one of central interest is rarely possible. We can, however, partial away the shared covariation of the dependent and independent variables with other explanatory factors, *conditional upon a theoretical model* like (5) that

[16] This is so to some extent even in the social science laboratory because the experimental subjects inherently bring with them potential explanatory factors the researcher cannot control. The researcher can randomize (or match) but even then can only evaluate empirically the efficacy of the randomization (and adequacy of the sample variation and size) across observables (see Przeworski, this volume, and Section 5 below).

explicitly states how these other explanatory factors relate to the outcome, thereby obtaining an estimate of the relationship between one factor and the outcome holding other factors constant even though we cannot and have not exactly held those other factors constant.[17] In other words, given that most everything varies most of the time, we can only gauge the partial association of any one independent variable with a dependent variable if we know how our independent variable covaries with other possible causes and how those others relate to the dependent variable. Furthermore, to net away the latter we must know, estimate, or be willing to stipulate some model of how controls (as well as independent variables of interest) relate to the dependent variable (see also n. 17). If *EcDev* and *EqDist* both relate to *Dem*, e.g., we cannot ascertain the effect of one without having some way of netting the effect of the other, regardless of whether we analyze the available information qualitatively or quantitatively.

Equation (6) shows explicitly how this "partialing away" of shared covariation is done in the simple linear, trivariate regression case of (5).[18] Jumping to the last lines of each open parenthetical section, we see that to gauge the partial association of one independent variable, say *EcDev*, with the dependent variable, *Dem*, controlling for the other, *EqDist*, one must assess the covariation of *EcDev* with dependent variable, *Dem*, net away the partial association of *Dem* with the other independent variable, *EqDist*, times the covariation of the two independent variables, *EcDev* and *EqDist*, and then divide by the variation in *EcDev*. (This appears toward the middle of equation (6).) Alternatively, if one cannot or prefers not to stipulate or assess a priori the partial association of the other independent variable, then one must multiply the covariation of *Dem* with *EcDev* by the variation of *EqDist*, subtract the covariation of *Dem* and *EqDist* times the covariation of *EcDev* and *EqDist*, and divide all that by the product of the variations of *EcDev* and *EqDist* minus their covariation squared.

$$Min_{b_1,b_2} \sum_{i=1}^{n} (Dem_i - b_1 EcDev_i - b_2 EqDist_i)^2$$

$$\Rightarrow \begin{cases} \text{(i)} & \dfrac{\partial \sum_{i=1}^{n} (Dem_i - b_1 EcDev_i - b_2 EqDist_i)^2}{\partial b_1} = 0 \\ \text{(ii)} & \dfrac{\partial \sum_{i=1}^{n} (Dem_i - b_1 EcDev_i - b_2 EqDist_i)^2}{\partial b_2} = 0 \end{cases}$$

[17] Propensity-score matching and related non- and semi-parametric techniques attempt to relax this need to leverage a particular specified empirical model of the relations between controls and outcomes in obtaining estimates for some particular variable's effect. They achieve this by replacing parametric model assumptions with others related, *inter alia*, to the distributions of the observed and unobserved controls (see Przeworski, this volume, and the discussion below). In either case, then, we can redress multi-causality in the social scientific context only by adding theoretical information. Again, social scientific experimentation, laboratory or, a fortiori, field, can only partially evade this need because it can only partially control possible causes.

[18] Indeed, we simplify yet further by assuming *EcDev* and *EqDist* each have sample mean zero so the last lines of each open parenthetical may be written in terms of sample variations and covariations.

$$
(\mathrm{i}) \Rightarrow \begin{cases} \sum_{i=1}^{n} EcDev_i(Dem_i - b_1 EcDev_i - b_2 EqDist_i) = 0 \Rightarrow \sum_{i=1}^{n} EcDev_i Dem_i \\ = \; b_1 \sum_{i=1}^{n} EcDev_i^2 + b_2 \sum_{i=1}^{n} EcDev_i EqDist_i \\ \Rightarrow b_1 = \left(\sum_{i=1}^{n} EcDev_i Dem_i - b_2 \sum_{i=1}^{n} EcDev_i EqDist_i \right) / \sum_{i=1}^{n} EcDev_i^2 \\ \Rightarrow b_1 = \dfrac{Cov(Dem, EcDev) - b_2 \times Cov(EcDev, EqDist)}{Var(EcDev)} \end{cases}
$$

$$
(\mathrm{ii}) \Rightarrow (\text{analogously}) \begin{cases} b_2 = \left(\sum_{i=1}^{n} EqDist_i Dem_i - b_1 \sum_{i=1}^{n} EcDev_i EqDist_i \right) / \sum_{i=1}^{n} EqDist_i^2 \\ b_2 = \dfrac{Cov(Dem, EqDist) - b_1 \times Cov(EcDev, EqDist)}{Var(EqDist)} \end{cases}
$$

$$
\Rightarrow \begin{cases} b_1 = \dfrac{\left(\sum_{i=1}^{n} EcDev_i Dem_i\right)\left(\sum_{i=1}^{n} EqDist_i^2\right) - \left(\sum_{i=1}^{n} EqDist_i Dem_i\right)\left(\sum_{i=1}^{n} EcDev_i EqDist_i\right)}{\left(\sum_{i=1}^{n} EcDev_i^2\right)\left(\sum_{i=1}^{n} EqDist_i^2\right) - \left(\sum_{i=1}^{n} EcDev_i EqDist_i\right)^2} \\ b_1 = \dfrac{Cov(Dem, EcDev) \times Var(EqDist) - Cov(Dem, EqDist) \times Cov(EcDev, EqDist)}{Var(EcDev) \times Var(EqDist) - [Cov(EcDev, EqDist)]^2} \end{cases}
$$

and, analogously:

$$
\Rightarrow \begin{cases} b_2 = \dfrac{\left(\sum_{i=1}^{n} EqDist_i Dem_i\right)\left(\sum_{i=1}^{n} EcDev_i^2\right) - \left(\sum_{i=1}^{n} EcDev_i Dem_i\right)\left(\sum_{i=1}^{n} EcDev_i EqDist_i\right)}{\left(\sum_{i=1}^{n} EcDev_i^2\right)\left(\sum_{i=1}^{n} EqDist_i^2\right) - \left(\sum_{i=1}^{n} EcDev_i EqDist_i\right)^2} \\ b_2 = \dfrac{Cov(Dem, EqDist) \times Var(EcDev) - Cov(Dem, EcDev) \times Cov(EcDev, EqDist)}{Var(EcDev) \times Var(EqDist) - [Cov(EcDev, EqDist)]^2} \end{cases} \tag{6}
$$

I remind the reader that this is the simplest possible case: purely linear additive, only two independent variables (each with sample mean zero), and an additively separable stochastic component. The expression becomes (exponentially) more complicated, not less, as we relax these extremely restrictive assumptions, for example by entertaining non-linear and/or interactive relationships or non-separable stochastic components (such as common in binary or other qualitative outcome models). Notice also that to gauge two parameters of interest, b_1 and b_2, we need at least three observations (i.e., to observe three contexts, three sets of information). This just reflects the obvious point that positive degrees of freedom are necessary to learn anything empirically (as elaborated next section).[19]

[19] Actually, in (6), just two will suffice to gauge the parameters because we stipulated rather than estimated the sample means of the variables, although to gauge their certainty would still require at least one more.

Thus, whether one works by qualitative or quantitative analysis, if more than one cause possibly operates and if each of those potentially varies across contexts considered, then one must gauge all of these quantities and perform the calculations in (6) at least loosely if one is to claim comprehension of the association of a variable with an outcome, controlling for others. Of course, qualitatively, working (implicitly) with Figure 2.1 would be preferred, but the figure is imperfect and not generally applicable as a representation of multivariate analysis and one must at least loosely gauge all of the above variations and covariations (or partial covarations) to draw it with appropriately sized and positioned ovals and overlaps anyway. Drawing such figures properly with more than two independent variables or for non-linear or other more complexly context-conditional cases is, to state it mildly, extremely difficult. We can clearly see from this exposition why such heavy premium is rightly placed in qualitative traditions on selecting contexts for close observation that literally fix as much as possible apart from a single or very, very few potential causes of interest. To state the conclusion more baldly and boldly, one simply cannot manage the complexity of partialing shared covariation in one's head, at least not easily or well, and certainly not with more than two moving potential causes, so, given multicausality, qualitative empirical analysts must (and rightly do in most cases) strive carefully and determinedly to isolate for analysis episodes from within their contexts in which only one potential cause at a time moves, preferably by a lot because only few instances of an uncertain effect will be observed; in short: seek contexts with big effects and single or very few moving causes.

We can also see from Figure 2.1 and equation (6) the intuitions behind the important and powerful omitted variable results from statistical analysis.[20] Suppose, for example, that, in addition to or instead of *EcDev* and *EqDist*, a cross-cutting rather than a reinforcing ethno-religio-linguistic social cleavage structure, *CCut*, fosters democratization. Some theories suggest, moreover, that *CCut* would also foster *EcDev* and *EqDist*. Accordingly, estimation of (5) or qualitative analysis of democratization that considered only *EcDev* and *EqDist* and did not or could not control for *CCut*, i.e., "partial away" its effects quantitatively or hold it fixed in qualitative analysis, would tend to overestimate the importance of *EcDev* and *EqDist*. We can see this most easily from the following line of equation (6):

[20] Figure 2.1 can also illustrate the oft-noted problem of multicolinearity. As the overlap of X_2's and X_3's ovals increases, gauging their partial relations with X_1 increasingly relies on mere slivers of unique covariation with X_1. Thus, multicolinearity induces greater uncertainty and larger standard errors, *but it does so correctly and without any associated bias*, in general. The researcher really is less certain of the association of X_2 with X_1 holding X_3 constant as X_2 and X_3 correlate more. The problem of multicolinearity is the uncertainty it correctly leaves about partial associations, not bias. It is an unfortunate fact of the information one has and not a failure of model specification or estimation strategy. Only more information, preferably new information in which the covariance of potential causes is lower, can help. On the other hand, this is one area where relatively greater emphasis on quality could also be very productive. With random and uncorrelated measurement error, those slivers of unique variation, and perhaps some of the covariation in limited samples, would comprise largely noise as measurement error and correlation among the true explanatory factors increased.

$$b_1 = \frac{Cov(Dem,\ EcDev) - b_2 \times Cov(EcDev,\ EqDist)}{Var(EcDev)}$$
$$= b_1^{\text{bivariate}} - b_2 \times b_{EqDist.on.EcDev} \tag{7}$$

The first term in the numerator divided by the denominator is the bivariate coefficient from a regression of *Dem* on *EcDev*. The term after the minus sign thus gives the bias in that bivariate coefficient relative to the corresponding partial coefficient from the (assumed correct) multivariate regression. The bias in the trivariate case is simply the (correct) partial coefficient on the omitted variable times the coefficient one would obtain regressing the omitted variable on the included one. Thus, the signs of omitted variable biases in the trivariate case are easily determined given some assumptions, argument, or theory about how the omitted factor relates to the dependent variable and the included independent variable.

The logic is intuitive: empirical analysis omitting some factors will attribute to factors that are included and that relate to the omitted ones a share of the associations of those omitted variables with the dependent variable proportional to the covariations of the omitted with the included variables. In our example, the omitted *CCut* was expected to relate positively both to the independent variables and to the dependent variable. If these former statements intended positive partial covariations with the included variables, then we would expect its omission to bias the researcher's conclusions about the effects of each *EcDev* and *EqDist* positively. However, if the statement intended only that *CCut* covaried with each positively, but that its partial covariation with each controlling for the other might be negative, this logic actually establishes only that the sum of the effects of the included, *EcDev* and *EqDist*, will be overestimated due to the omission of the third, *CCut*.

The general formula for omitted variable bias in the *k*-regressor multivariate case is:

$$\left[(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right]\mathbf{b}_2 \equiv \mathbf{F}_{1,2}\mathbf{b}_2 \tag{8}$$

Table 2.2 Sign of omitted-variable bias from bivariate analysis of trivariate relationship

	Sign of partial dependent on	coefficient on omitted independent variables	factor from trivariate regression of	
Sign of coefficient from regression of omitted on Included variable		Positive	Zero	Negative
	Positive	Positive Bias	No Bias	Negative Bias
	Zero	No Bias	No Bias	No Bias
	Negative	Negative Bias	No Bias	Positive Bias

where $\mathbf{F}_{1,2}$ is a $k_1 \times k_2$ matrix of partial coefficients obtained from regressing the vector of k_2 omitted factors on the vector of k_1 included factors and $\mathbf{b}_2$ is the vector of partial coefficients on those omitted factors. Thus, in our example, $\mathbf{b}_2$ is just the single partial coefficient on the omitted *CCut*, which was assumed positive, and $\mathbf{F}_{1,2}$ is the vector of two partial coefficients obtained from regressing *CCut* on *EcDev* and *EqDist*. Thus, if partial coefficients on *EcDev* were positive and on *EqDist* negative, then the association of *EcDev* with *Dem* would be overstated and that of *EqDist* with *Dem* understated if *CCut* were omitted or ignored.

The potential biases from omitted variables, and the inherent difficulty discussed above in gauging the partial association of multiple included and omitted variables, represent the first of the fundamental challenges for empirical evaluation in the complex, multicausal world of social science. As readily noted from Table 2.2 and equations (6) and (8), the signs and magnitudes of omitted variable biases are relatively easily determined and gauged, qualitatively or quantitatively, in the trivariate case, but they grow exponentially more complicated to assess, especially qualitatively, as the number of potentially important causal factors grows. Moreover, if information on the omitted potentially relevant factors can be obtained, then including them seems at first blush relatively costless for quantitative analyses. (For qualitative empirical analysis, the difficulty of partialing the associations due even just two moving explanatory factors already suffices to place extremely high premium on choosing contexts across which just one or as few as possible potential explanatory variables vary.) Return to equation (7) or Table 2.2, and notice that if the hypothetical omitted factor in them, *EqDist*, were actually irrelevant to the dependent variable, then the middle column of Table 2.2, the case where $b_2 = 0$ in (7), would apply and bias from including or excluding such an irrelevant factor is zero.[21] This line of reasoning has generally led to a predisposition among quantitative empirical researchers "to err on the side of caution" and include any and all reasonably plausible controls in their estimation models.

The seemingly cautious strategy of considering and controlling many factors, however, has its own serious perils. First is *overfitting*. When one includes too many explanatory factors (or too flexibly includes some number), then one tends to find in limited samples associations of those independent variables with the random component that happens to have realized in that sample. Oversized models in this sense do correspondingly poor jobs of out-of-sample prediction.[22] Second is what might be termed *included variable bias*. In short, control should be applied only for causal priors and not causal posteriors. If, for example, economic development

[21] This assumes the included irrelevant (in this sense) factor is exogenous. If endogenous, then its relationship with the dependent variable may be misgauged (as non-zero), inducing biases in other variables' coefficient estimates (see, e.g., Franzese and Hays 2006).

[22] Many or most quantitative empirical analyses are likely overfit thusly. Many or most qualitative empirical analyses probably are as well, although for different reasons. Namely, when describing in qualitative detail events and circumstances in a limited set of contexts, one often feels almost compelled to explain everything about those contexts. If the sociopoliticoeconomic world is partly random, then complete explanation in this sense necessarily includes erroneously systematic seeming accounts of non-systematic (i.e., random) aspects of those events and circumstances. The caution against overfitting therefore, like almost everything discussed in this chapter, applies to quantitative and qualitative empirical analyses alike.

affects democratization through the effect of development on equality, then controlling for equality will induce understatement of development's democratization effect. In Figure 2.1, the areas [3]+[6] reflect X_2's and X_3's shared covariation, but steem solely from variation in X_2, and so should not be partialed away if one seeks to gauge the total impact of X_2 on X_1.[23] Third, even quantitatively, the complexity with which various sources of incorrect inference tend to interact in multicausal models, especially if we add mismeasurement or misspecification possibilities, should lead researchers to place great weight on parsimony. After discussing some of these important and complex difficulties, Achen (2002) suggests restraining research questions to a narrower range of contexts, leaning more heavily on theory to help specify empirical models and explorations, and, famously, "ART: a rule of [no more than] three [explanatory factors]." Unfortunately, the first piece of advice is not applicable for comparativists who seek empirically helpful general theory rather than an unconnected set of partial theories that may be empirically helpful each in their narrow context.[24] The third, as stated in the catchy ART, is clearly not helpful if understood too simplistically and rigidly (which was not the intent). The problems and complexities Achen discusses are real and very important, but, unfortunately, so are the problems of omitted variables, even though Achen is also correct to argue that the omitted variable problems are often overemphasized to the exclusion of the equally difficult problems he stresses. Parsimony is certainly to be valued, and Achen's second admonition, that we rely more heavily and directly than currently common on our theoretical models and arguments (and substantive knowledge) to specify our empirical models and explorations, is certainly to be followed, but no rigid rules or limits will ever suffice to encapsulate those valuable guides, and three will often prove too few, sometimes way too few, to capture even just the most important potential causes.[25] Unhelpful as the following may be, the only general

[23] Notice, however, that, as the example is drawn, simply omitting X_3 would be inefficient—more exactly, the estimated variance of the estimated coefficient on X_2 would be higher than it could (correctly) be—because some share of the dependent variable, X_1, would be treated as stochastic (random) whereas in fact, it is systematic in X_3. In such a case, we might wish to include $X_3|X_2$, i.e., X_3 net of X_2, as a regressor. Notice then, too, that this would stipulate a priori that all shared covariation of X_2 and X_3 with X_1 is attributed to X_2. In the more complicated (and probably more common) case where we are unable or do not wish to make such stipulation, we are back in the original multicausal case where no empirical information, quantitative or qualitative, can determine how to allocate the shared covariation. Any specific allocation can only be imposed a priori by theory or assumption. The interested reader should reference a good text on structural equations, e.g. Duncan 1975, however because the issues involved are more complex than this note can fully relay.

[24] See also the discussion in Section 2 about the empirical uses of and strategies for "single-case" studies. Moreover, paradoxically, the latter "empirical [helpfulness] each in their narrow context" may often prove difficult to adjudge with information only from that narrow context, especially given multicausality. One problem is that, typically, a great many potential causes will be held constant or not vary much by the narrowed focus; accordingly, if they are indeed causes, means of gauging their impact relative to those that do vary within the narrow context will not exist. Again, see also the discussion of "single cases" in Section 2.

[25] Huntington's discovery of twenty-seven in the democratization literature is probably too many, but illustrative. A model of democratization that omitted history (the previous state of the regime), economic development, social structure, or international conditions would almost certainly be badly misspecified, for example (and the important aspects of each of those, especially social structure, almost certainly number more than one).

advice one can offer on the number of factors to include in empirical models and the proper way to specify their inclusion is "the right number, the right way." The importance and full meaning of the latter part of that banality, attention to empirical specification, will become manifest as we turn now to the second of the fundamental challenges for empirical evaluation in social science, the one most central to comparative politics' core tenet that *context matters*: the effect of everything depends on pretty much everything else.

4. Context-Conditionality: The Effect of Almost Everything Depends on Almost Everything Else

The contention that *context matters* perhaps means most centrally that how particular causes operate, and perhaps the entire structure of the causal process, is highly contextual; the causal process varies across contexts: it is context conditional. Such contentions and the concerns they raised regarding the prospects for successful general empirical evaluation of theoretical propositions may have underlain the parochialism and non-comparativeness of the pre-war comparative politics that is generally viewed (probably unfairly) by all modern comparativists as pre-scientific. In the extreme, *context matters* means that processes and outcomes differ uniquely, each specific substantive venue in each place at each time having its own unique processes relating to outcomes. If this is so, *any* comparison, within or across cases, times, or venues, would always be unwarranted or unhelpful. Under these conditions, as I show formally below, one simply could not and cannot learn any more than description from comparison, history, or comparative history. Some scholars may have realized this, and some may even have appreciated and intended it, but many seemed (and some seem) to think one could hold simultaneously that each context was unique in this sense and that one could learn from comparative history. That is not logically possible.

Such contentions and concerns were also foundational to early cultural and behavioral approaches to comparative politics, in which the meaning and effect of various objective circumstances and factors (e.g., deprivation) depend on cultural and sociopsychological context (e.g., perceived or relative deprivation). However, with this wave of theoretical progress, contextual variation became something to be understood better by comparison, rather than something debarred by it. Likewise in institutional approaches, the effects of societal interest structures *manifest through*, are *shaped by*, and therefore *depend upon* the institutional structure of the society, economy, and polity. Institutions, in other words, became a key to understanding context-conditionality. At least by the 1980s, perhaps all comparativists could agree, scholars had also established that statistical analysis of comparative-historical data

could help evaluate and fruitfully inform positive theories relating social structure, culture, and institutions to political outcomes. The early empirical work, quantitative and qualitative alike, on these two theoretical approaches did not often reflect as well and fully as one might like the multicausality, and rarely reflected at all the context-conditionality and endogeneity, of comparative-politics arguments and reality. For example, culture matters, if it does, in complex ways and often by modifying the relations between other objective conditions, like poverty and underdevelopment, and outcomes, like democratic stability. Individuals' interpretation of poverty and appropriate responses thereto, a cultural argument might argue, hinge on cultural symbols and understandings. Likewise, institutions matter mostly by altering the relationships between objective interests and the institutionally shaped set of actions perceived as possible and most effective by individuals or groups with those interests. For example, the extent to which some polity's cleavage structure will induce (i) leaders to form political parties mirroring the societal groups drawn by that structure and (ii) voters to support such parties depends upon the electoral rules and party-systemic strategic structure that shape the relationships from votes to representation and from representation and governmental power.

Complex, context-conditional propositions of this sort are now the hallmark of positive comparative politics; the effect of X (e.g., institutions) on Y (outcomes) *depends on* Z (culture, structure, etc.): formally, $\partial Y/\partial X = f(Z)$. Early empirical work that established that institutions matter *in addition to* culture and structure (and vice versa) by controlling for the latter in regressions of outcomes on the former, reflected multicausality, and so faced the challenges to empirical inference thereof as discussed above, but they did not reflect such context-conditionality. They showed only that the effect of X (institutions) on Y (outcomes), given or controlling for Z (culture, structure, etc.) is not zero, not that the effect of X on Y *depends on* Z: formally, they showed $\partial Y/\partial X|_Z \neq 0$ and not (necessarily) that $\partial Y/\partial X = f(Z)$.

Critics of statistical analysis in comparative politics often cite this concern (*inter alia*) that regression coefficients impose a *constant* effect for each independent variable, albeit controlling for others, not effects that differ depending on context. That is, they object that broad statistical comparison neglects the context-conditionality that lies at the very heart of comparative politics. This criticism, however, applies only to the simplest linear-additive regression. Nor does it follow that other approaches necessarily evade this or any other limitation simply because one has discovered or claims a weakness in one approach. Hall (2003), e.g., offers perhaps the most careful, nuanced, and best statement of this concern (among others to which the next section returns):

> Regression analysis is more flexible. It is well-adapted to an ontology that envisions probabalistic causation and, given enough cases, it can cope with some interaction effects (cf. Jackson 1996). However, the types of regression analyses commonly used to study comparative politics . . . assume unit homogeneity, which is to say that, other things being equal, a change in the value of a causal variable x will produce a corresponding change in the value of the outcome variable y of the same magnitude across all the cases. It assumes no systematic correlation between the causal variables included in the analysis and other causal variables

omitted from it. It assumes that all the relevant interaction effects among the causal variables have been captured by interaction terms in the regression . . . [26]

As Hall notes, the statistical device most frequently used to evaluate theoretical claims that the effect(s) on some dependent variable(s), *Y*, of some independent variable(s), *X*, depend upon or are moderated by a third (set of) independent variable(s), *Z*, is the linear-interactive, or multiplicative, term. One simply includes one or more $X \times Z$ terms among the regressors. Interaction terms are hardly new to comparative politics. Indeed, their use is now common, yet, especially with current and growing attention to the roles of institutions in comparative politics, they should perhaps become more common still. Moreover, as we will elaborate later, many statistical devices exist to incorporate the context-conditionality of comparative phenomena (of any complexity) into empirical models. In fact, if one can make a logically consistent claim that theory predicts some relationship between *Y* and *X* (and chance), $Y = f(X, \epsilon)$, then one can write a statistical model to reflect that proposition, and, if the necessary empirical information actually exists, estimate and evaluate it. And, again, the challenges to redress in doing so are not a function of the empirical methodology chosen, but rather logically inherent in the attempt to infer complex context-conditionality from comparative history. They do not arise just because we write the problem formally, and they do not disappear just because we do not.

As Table 2.3 (from Kam and Franzese forthcoming) shows, 54 percent of articles 1996–2001 in leading political science journals use some statistical methods,[27] and 24 percent of those employ interactions. Among exclusively comparative journals, *Comparative Political Studies* had 49 percent and 25 percent and *Comparative Politics* 9 percent and 8 percent. The other journals all have many comparative publications also, and statistical analyses comprise 25–80 percent of those articles, with interactive analyses representing a relatively fixed 20–5 percent. Thus, about half of top journal political science articles employ some statistical methods, and about a quarter of those and over an eighth of all articles use interaction terms.[28] Comparative politics, using *CPS*, *IO*, and *WP* to gauge that, operates somewhere between the discipline's mean and half that on these dimensions. Trends in comparative politics and the broader discipline likely continue mildly upward in both regards. Notwithstanding this widespread and expanding usage of interactions, still more empirical work should contain them than currently does, given the substance of many comparative politics arguments. Consider the gist of most institutional arguments, for example. In one influential statement, Hall (1986, 19) states:

[26] The last two sentences in the quotation refer to the multicausality and omitted variable bias discussed in Section 3; we need re-emphasize here only that the potential problem is not one of regression analysis but of empirical evaluation of social science. Qualitatively or quantitatively, valid empirical inference rests on assumptions or arguments that one has controlled for or has randomized over other potential causes (and has observed sufficient and sufficiently independent information sets for randomization to be effective). We have already discussed also the relative efficacy of quantity and quality in making such control or randomization.

[27] i.e., they report some certainty estimate(s) like standard errors, confidence intervals, or hypothesis tests.

[28] Moreover, the denominator includes formal theory and political philosophy, and implicitly interactive functional forms, like logit or probit, are not counted in the numerator. These are very conservative estimates.

Table 2.3 Types of articles in major political science journals, 1996–2001

Journal (1996–2001)	Total articles	Statistical analysis		Interaction-term usage		
		Count	% of tot	Count	% of tot	% of stat
American Political Science Review	279	274	77	69	19	25
American Journal of Political Science	355	155	55	47	17	30
Comparative Politics	130	12	9	1	1	8
Comparative Political Studies	189	92	49	23	12	25
International Organization	170	43	25	9	5	21
International Studies Quarterly	173	70	40	10	6	14
Journal of Politics	284	226	80	55	19	24
Legislative Studies Quarterly	157	104	66	19	12	18
World Politics	116	28	24	6	5	25
TOTALS	**2,446**	**1323**	**54**	**311**	**13**	**24**

institutional analysis of politics . . . emphasizes institutional relationships, both formal and conventional, that bind the components of the state together and structure its relations with society . . . [I]nstitutions . . . refers to the formal rules, compliance procedures, and standard operating practices that structure the relationship between individuals in various units of the polity and economy . . . Institutional factors play two fundamental roles . . . [They] affect the degree of power that any one set of actors has over policy outcomes [. . . and they . . .] influence an actor's definition of his own interests, by establishing his . . . responsibilities and relationship to other actors . . . With an institutionalist model we can see policy as more than the sum of countervailing pressure from social groups. That pressure is mediated by an organizational [i.e., institutional] dynamic . . . (emphases added)

Thus, in this approach to institutional analysis, and, as we argued above, inherently in all approaches, institutions are intervening variables that *funnel*, *mediate*, or otherwise *shape* the political processes that translate the societal structure of interests into effective political pressures, those pressures into public policy-making responses, and/or those policies into outcomes.[29] For example, one prominent line of research connects the societal structure of interests to effective political pressure on policy makers through institutional features of the electoral system: plurality-majority versus proportional representation, etc. (e.g., Cox 1997; Lijphart 1994). Another emphasizes how governmental institutions, especially the number and polarization of key policy makers (veto actors) that comprise it, shape policy-making responses to such pressures (e.g., Tsebelis 2002). A third stresses how the institutional configuration of the economy, such as the coordination of wage–price bargaining,

[29] Extending the list of synonyms might prove a useful means of identifying interactive arguments. When one says *X alters, modifies, magnifies, augments, increases, moderates, dampens, diminishes, reduces,* etc. some *effect* (of *Z*) on *Y*, one has offered an interactive argument.

shapes the effect of certain policies, such as monetary policy (e.g., Franzese 2002b, ch. 4). In every case, and at each step of the analysis from interest structure to outcomes (and back), the role of institutions is to mediate, shape, structure, or condition the impact of some other variable(s) on the dependent variable of interest. That is, institutional arguments are inherently interactive, yet, with relatively rare exceptions—see, e.g., Ordeshook and Shvetsova 1994; Franzese 2002b, ch. 3; Franzese 2002b, ch. 4, respectively, regarding the above topical examples—empirical work on institutional arguments has ignored this interactivity.

Another example further reveals the ubiquity of the interactive, i.e., context-conditional, implications of comparative-institutional theories. Scholars consider principal–agent (i.e., delegation) situations interesting, problematic, and worthy of study because, if each had full control, agents would determine policy, y_1, in response to some (set of) factor(s), $\boldsymbol{X}$, according to some function, $y_1 = f(\boldsymbol{X})$, whereas principals would respond to some perhaps different (set of) factor(s), $\boldsymbol{Z}$, perhaps differently according to, $y_2 = g(\boldsymbol{Z})$. (For example, the principals might be the current government, responding to political-economic conditions $\boldsymbol{X}$, and the agents unresponsive central banks, giving $\boldsymbol{Z}=\emptyset$, as in Franzese 1999.) Theorists then offer some arguments about how institutional and other environmental conditions determine the monitoring, enforcement, and other costs, $\boldsymbol{C}$, principals must incur to induce agents to enact $g(\boldsymbol{Z})$ instead of $f(\boldsymbol{X})$. In such situations, realized policy, y, will usually be given by some $y = k(\boldsymbol{C}) \cdot f(\boldsymbol{X}) + [1-k(\boldsymbol{C})] \cdot g(\boldsymbol{Z})$ with $0 \leq k(\boldsymbol{C}) \leq 1$ and $k(\boldsymbol{C})$ weakly increasing. Thus, the effects on y of each $c \in \boldsymbol{C}$ generally depends on $\boldsymbol{X}$ and $\boldsymbol{Z}$, and those of each $x \in \boldsymbol{X}$ and $z \in \boldsymbol{Z}$ generally depend on $\boldsymbol{C}$. That is, the effect on y of *everything* that contributes to monitoring and enforcement costs generally depends on *all* factors to which the principals and agents would respond differently, and, vice versa, the effect on y of *all* such factors depends on *everything* that affects monitoring and enforcement costs. In fact, policies and outcomes in all situations of shared policy control, through delegation or otherwise (e.g., presidents and legislatures), will usually be describable as convex combinations like these, implying the corresponding multiple and complex interactions. Most empirical applications of principal–agent and other shared policy control models seem to have missed this point.

A rough quantification of the magnitude of such institutional interactions omissions from empirical specifications is startling. Of Table 2.3's 1,012 articles with non-interactive statistical analyses, half or so offer some sort of institutional argument. Even if only half of all institutional effects actually reflect the interactivity argued here to be inherent, then almost as many articles, $\frac{1}{2} \cdot \frac{1}{2} \cdot 1012 = 253$, incorrectly employ non-interactive empirical techniques to evaluate interactive hypotheses as actually employ interactive terms (311). If, instead, all institutional arguments are inherently interactive, and many other arguments are also (e.g., contextual effects in cultural-behavioral theories), say half, that would imply that roughly 2.5 times as many articles made interactive arguments but empirically evaluated them non-interactively ($\{1/2 + 1/4\} \cdot 1012 = 759$) as actually employed interactions.

The theoretical and substantive interestingness of such complex context-conditionality is readily apparent in comparative political economy. For example, in a

recent review, Franzese (2002a) argued that venerable electoral and partisan cycles may merit theoretical and empirical revisit to explore the institutional, structural, and strategic conditionality:

> Policymakers in democracies have strong partisan and electoral incentives regarding the amount, nature, and timing of economic-policy activity. Given these incentives, many observers expected government control of effective economic policies to induce clear economic-outcome cycles that track the electoral calendar in timing and incumbent partisanship in character . . . until recently, both rational- and adaptive-expectations electoral-and-partisan-cycle work underemphasized crucial variation in the contexts—international and domestic, political and economic, institutional, structural, and strategic—in which elected partisan incumbents make policy. This contextual variation conditions policymaker incentives and abilities to manipulate economic policy for electoral and partisan gain, as well as the effectiveness of such manipulation, differently across democracies, elections, and policies. Although relatively new, research into such context-conditional electoral and partisan cycles seems to offer much promise for resolving anomalies and an ideal substantive venue for theoretical and empirical advancement in the study of political economy and comparative democratic politics more generally. (p. 369)

For example, in small, open economies, domestic policy makers may retain less autonomy over some policies, or some policies may be less economically effective, so that electoral and partisan cycles in those policies and outcomes are less pronounced than in larger, less-exposed economies. Some polities, moreover, concentrate policy-making control in fewer, more disciplined partisan actors, which may induce sharper cycles in, e.g., Westminsterian than in other democracies. Furthermore, some policies may have more effect and so be more useful and so more used for electoral or partisan purposes, and this too varies with institutional, structural, or strategic context. For instance, political benefits of geographic relative to demographic targeting of spending may vary by electoral system, e.g., single-member plurality favoring the latter and proportional representation the former. These and other contextual variations condition policy makers' incentives and abilities to manipulate policies and outcomes for electoral and partisan gain, and modify the political and economic efficacy of such manipulation, in manifold ways across democracies, elections, and policies, all of which suggests exciting opportunities for interactive models that inform comparative politics. Another obvious locus of interactive effects lies in recent studies of *Varieties of Capitalism* (Hall and Soskice 2001) or of globalization, the comparative political economy approach to which stresses that the domestic response to international economic integration varies, depending critically on domestic political and institutional context (e.g., Boix 1998; Garrett 1998; Swank 2002). Similar examples from outside political economy are not hard to imagine. The propensity for (apparent) directional voting versus proximity voting in individual electoral behavior, for example, depends on electoral and party systems and the types of government they tend to produce (see, e.g., Kedar 2002).

With so many opportunities, some currently being taken but many as yet ignored, to explore interactions—indeed, with the logically inherent interactive nature of comparative politics theory—the good news is that quantitative empirical modeling of such context-conditionality can be quite simple (Brambor, Clark, and Golder 2006 and Kam and Franzese forthcoming, discuss more thoroughly). First, one must

understand empirical models that embody interactive hypotheses. For example, one typical theoretical argument might be that X generally reduces Y and does so more in the presence of or the larger is Z. Note that this is actually *two* hypotheses: (a) that $\partial Y/\partial X$ is negative (X reduces Y) and (b) that $\partial^2 Y/\partial X \partial Z$ is negative (and increasingly so with Z). In a model containing regressors X, Z, and $X \times Z$, such as $Y = \ldots aX + bZ + cX \times Z \ldots$, the interpretation of results regarding (b) is straightforward. $\partial^2 Y/\partial X \partial Z = c$, so the coefficient c simply and directly tells us how the effect of X changes per unit increase in Z and, vice versa, how the effect of Z changes per unit increase in X.[30] Thus, the standard t-test on coefficient c corresponds to hypothesis (b). The effect on Y of X, $\partial Y/\partial X$, however, is not simply a, nor is the effect on Y of Z, $\partial Y/\partial Z$, equal to b; nor, even, are these the "main" effects of X or Z. The effect of X on Y, $\partial Y/\partial Z$, equals $a + cZ$, which depends, as the hypothesis said, on the value of Z (and vice versa: $\partial Y/\partial Z = b + cX$). The effects of X and Z each depend on the other variable's value, and the coefficient a or b is just the effect of an increase in X or Z when the other variable equals zero (which need not be "main" in any way, and could even be out of sample or logically impossible). In interactive models, as in any models beyond the strictly linear additive, the *effects* of variables do not correspond directly to just one *coefficient*; effects of each variable depend on the values of their interacting variables, which is what the interactive argument argued in the first place. Nor do the standard errors (or t-tests) of these *effects* correspond directly to those of any one coefficient; just as the effects of X and Z vary depending on the value of the other, so too do the standard errors of those effects. The best approach for researchers presenting interactive results is to graph or tabulate the *effect* of each variable as a function of the value of its interacting variables, along with the standard errors or confidence intervals of those effects. Even with a relatively firm understanding of interactive models, some scholars express considerable reservations over them or question how far they can go toward reflecting and evaluating the complex context-conditionality of comparative politics.

Some note, correctly, that the empirical task of distinguishing not just a single, constant effect for X, but one that varies (albeit only linearly) depending on Z, imposes much heavier burden on the data. This is also the substantive meaning of concerns expressed regarding the high multicolinearity (i.e., correlation) among regressors X or Z and $X \times Z$ in interactive models.[31] Efficiency (but not bias) concerns over multicolinearity are quite valid, as we discussed already above. The empirical

[30] These converses are logically identical, and this identity is logical, not a result of regression modeling.

[31] Moreover, one must discard the notion that "centering" the interacting variables (subtracting their means), as some methodological texts advise, eases this empirical task. *Centering alters nothing important mathematically and nothing at all substantively.* Likewise, the oft-raised concern that multiplicative terms cannot distinguish, for example, $XZ=12$ with $X=3$ and $Z=4$ from $XZ=12$ with $X=2$ and $Z=6$, is incorrect because the model, that is the model of the effect of X and Z on Y, can and will distinguish those cases insofar as they actually do differ logically. Incidentally, under the heading of potentially misleading common admonitions, that one should include both X and Z if the model contains an XZ term is usually a highly advisable philosophy of science guideline (Occam's razor), and typically soundly cautious and conservative scientific practice at least to explore, but it is neither a logical nor a statistical necessity (see Kam and Franzese forthcoming).

task that interactive analyses set *is* very demanding, and these demands will heighten sharply as the number and complexity of interactions increase, as the complex context-conditionality of comparative politics suggests they should. However, this concern too is unavoidable *logical necessity*, and not a function of the empirical methodology chosen. Indeed, the difficulty of the task increases with the number and complexity of the interactions *relative to the number of—more exactly, the useful total variation in—the sets of information used to evaluate them.*

Comparative researchers seem to have four options, each with characteristic perils. (i) Ignore the context-conditionality of their arguments by omitting interactive terms. Judging by Table 2.3, most analysts do this, but this does violence to the inherently (and interestingly) interactive nature of comparative politics and plagues those effects actually estimated with omitted variable bias and inefficiency. (ii) Reduce context-conditionality by allowing only one or few of the hypothesized interactions in their empirical model. This enables more exclusive focus on those included interactions and reduces the omitted variable biases and inefficiencies relative to excluding them altogether, but it does not eliminate these problems and it ignores the likely complexity of the context-conditionality in comparative politics. (iii) Constrain the empirical model of context-conditionality to follow a specific functional form suggested by theory (see, e.g., the above regarding principal–agent models; Franzese 1999, 2002c). This reduces the empirical-inferential demands on the data to reveal more of the theorized complex, context-conditionality in comparative phenomena, thereby reducing further the misspecification and inefficiency issues of the previous approaches, but many comparative theories may not be sufficiently precise to determine the form of interactions, the gained strength arises from leaning more heavily on theory, and the multicolinearity concerns re-emerge, albeit at a lesser pace, as the allowed complexity increases. (iv) Conduct closer (i.e., qualitative) analysis (of fewer contexts) to supplement or substitute for quantitative analysis. This may partly counteract the information deficiency that is the multicolinearity problem by enriching the detail and depth of the empirics, but it typically enhances the quality of the information thusly at the cost of severely reducing the quantity, a tradeoff Section 2 showed is unfavorable in many circumstances. Furthermore, the ability to discern complex interactions qualitatively, i.e., without precise numerical measurement and statistical control of independent variables, is inherently more difficult. Indeed, Section 3 showed that qualitative analysis of contexts in which more than one or two potential explanatory factors vary was exceedingly difficult already, without adding interactive context-conditionality, and even a single linear interaction will generally require variation in three explanatory factors, X, Z, and $X \times Z$, and so at least four information sets.

The third of these options, therefore, seems most promising. Ultimately, the problems raised by complex context-conditionality are logically inherent, so qualitative recourse cannot evade them and the other two options evade them only to the degree they suppress the (interesting) conditionality. To see the promise of this third approach, return to the principal–agent (i.e., delegation) situation described above. Generally, in such situations, we argue that, if each had full control, agents would act according to some function, $y_1 = f(\boldsymbol{X})$, while principals would act differently,

$y_2 = g(\boldsymbol{Z})$. We then argue that some institutional and other contextual conditions determine the monitoring, enforcement, and other costs, $\boldsymbol{C}$, principals must incur to force agents to enact $g(\boldsymbol{Z})$ instead of $f(\boldsymbol{X})$. Realized policy, y, will then typically be given by some $y = k(\boldsymbol{C}) \cdot f(\boldsymbol{X}) + [1 - k(\boldsymbol{C})] \cdot g(\boldsymbol{Z})$ with $0 \leq k(\boldsymbol{C}) \leq 1$ and $k(\boldsymbol{C})$ weakly increasing as noted. If the comparative theory can identify $k(\boldsymbol{C})$, that is, the function $k(\cdot)$ and contextual conditions, $\boldsymbol{C}$, that determine the degree to which principal or agent has effective control, and the functions $f(\cdot)$ and $g(\cdot)$ and factors $\boldsymbol{X}$ and $\boldsymbol{Z}$ that state to what and how principal and agent would respond if wholly (hypothetically) in charge, and if these functions and/or factors are not identical, then non-linear regression techniques (as, e.g., in Franzese 1999, 2002c) can gain leverage on *all* the complex conditional effects predicted in that comparative context. Moving beyond delegation to other situations of shared policy control, researchers might also fruitfully apply this approach to study the relative weight in policy control of, e.g., executive and legislative branches in (semi-)presidential systems, or of different chambers in multicameral systems, or of prime, cabinet average, cabinet median, and portfolio ministers in parliamentary systems, or of committees or cabinets and legislature floors or backbenchers or oppositions, or, even, of the degree to which elected representatives act legislatively as if they represent the residents of their electoral district, those therein who support them, or their national party constituency. Finally, even more generally, researchers can apply similar non-linear approaches to any situation in which some factor or set of factors modify the impact of several others proportionately, thereby bringing many more of their highly interactive theoretical propositions under empirical scrutiny than perhaps previously thought possible. Indeed, institutions often operate in this way. For example, institutions that foster greater party discipline may induce legislators to behave less geographically distributively and more class/ideological redistributively, implying a proportionate modification in their response to a range of political-economic conditions. Similarly, institutions that facilitate voter participation tend to broaden the distribution of interests represented in the electorate (and so in policy), again suggesting that such electoral institutions will proportionately modify the effect of many political-economic conditions on government policies (see, e.g., Franzese 2002, ch. 2).

Non-linear regression is simple to describe, given an understanding of linear regression. As noted above, the empirical implications of positive theory will usually amount to some statement that an outcome, y, depends on random chance, , and some explanatory factors, $\mathbf{x}$, perhaps including multiplicative interactions or other complex terms, according to some function, $y = f(\mathbf{x}, \boldsymbol{\beta}, \epsilon)$, involving parameters $\boldsymbol{\beta}$ that relate $\mathbf{x}$ to y. In linear regression, we assume the function is linear additive and separable, with $\boldsymbol{\beta}$ being simple coefficients on $\mathbf{x}$, giving $y = \mathbf{x}\boldsymbol{\beta} + \epsilon$. The ordinary linear regression problem and solution is thus:

$$
\begin{aligned}
Min_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i\boldsymbol{\beta})^2 = Min_{\beta}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \; & Min_{\beta}\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
& \Rightarrow \frac{\partial(\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = 0 \Rightarrow -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0 \\
& \Rightarrow \mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
& \Rightarrow \hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (9)
\end{aligned}
$$

If we instead continue to assume the random component is additively separable but allow explanatory factors, $\mathbf{x}$, and associated parameters, $\boldsymbol{\beta}$, to determine the systematic component of y according to some nonlinear function, $E(y) = f(\mathbf{x}, \boldsymbol{\beta})$, specified by theory, we have the following non-linear regression problem and solution:

$$
\begin{aligned}
Min_{\beta}(\mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta}))'(\mathbf{y} - f(\mathbf{X}, \boldsymbol{\beta})) = Min_{\beta}\, \mathrm{S} \\
& \equiv \mathbf{y}'\mathbf{y} - \mathbf{y}'f(\mathbf{X}, \boldsymbol{\beta}) - f(\mathbf{X}, \boldsymbol{\beta})'\mathbf{y} + f(\mathbf{X}, \boldsymbol{\beta})'f(\mathbf{X}, \boldsymbol{\beta}) \\
& \Rightarrow \frac{\partial S}{\partial\boldsymbol{\beta}} = 0 = -2\left(\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)'\mathbf{y} + 2\left(\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)'f(\mathbf{X}, \boldsymbol{\beta}) \\
& \Rightarrow \left(\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)'\mathbf{y} = \left(\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)'f(\mathbf{X}, \boldsymbol{\beta}) \qquad (10)
\end{aligned}
$$

If $f(\mathbf{x}, \boldsymbol{\beta})$ is the linear-additive $\mathbf{x}\boldsymbol{\beta}$ as in the ordinary regression problem, then the last expression solves analytically to the familiar OLS formula in (9). However, if $f(\mathbf{x}, \boldsymbol{\beta})$ is non-linear in parameters $\boldsymbol{\beta}$, then the last expression in (10) cannot in general be simplified further. $\hat{\boldsymbol{\beta}}_{NLS}$ may be found numerically (i.e., computer search) though, either by finding the values for $\boldsymbol{\beta}$ that satisfy that last expression or by finding the values that minimize the sum of squared errors, S, given the data, $\mathbf{y}$ and $\mathbf{X}$. Effectively, the *derivatives*[32] of $f(\mathbf{x}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, which are just $\mathbf{x}$ in the linear-additive case, serve as the regressors (and play a like role in estimating the variance of the estimated parameters). In short, our basic understandings about ordinary least squares regression, its necessary assumptions, and its properties under those assumptions, applies to non-linear regression with the *derivatives* of $f(\mathbf{x}, \boldsymbol{\beta})$ replacing $\mathbf{x}$.[33] The crucial change lies in interpretation and is the one that comes with any move beyond strictly linear-additive models—even just to simple linear interaction models, dynamic models (i.e., models with time or spatial lags of the dependent variable in them), or the familiar logit or probit models of (probabilities of) binary outcomes—namely, that *coefficients* are not *effects*. The effect of X on Y is, always and everywhere, the derivative or difference of (change in) Y with respect to (over the change in) X,

[32] Actually, the correct term is *gradient* because $\boldsymbol{\beta}$ is a vector, so the slope is multidimensional.

[33] All the usual additional complications of numerical optimization as opposed to analytical solution—such as possibility of local maxima, flat areas or ridges, or "nasty" surfaces to search and the concomitant need to explore multiple starting values and search sensitivities and procedures—apply also.

dY/dX, but only in purely linear-additive-separable models are these effects, these derivatives, equal to the coefficient on the variable in question. In other models, effects of one variable generally depend on other variables' values and usually more than one coefficient—that is, the effects of *X* are context conditional. The important point here is that, if we can theorize how *Y* depends on *X* with logical consistency, then we can write a function[34] that describes that relationship, and then we can specify our empirical model by that function. Finally, provided the specified equation is identified and has positive degrees of freedom so that empirical evaluation from comparative history is logically possible, and if comparative history has actually given us sufficient useful variation, we can estimate, evaluate, and interpret that model. In other words, complexity hardly debars statistical empirical analysis; in fact, as the discussions throughout this chapter suggest, complexity tends rather to argue strongly for such analysis. (Furthermore, the statistical software packages that political scientists commonly use now possess user-friendly NLS procedures.[35])

The approach is not magic, of course; it does have prerequisites and limitations. Most importantly, researchers must have sufficiently precise theory to specify empirical models usefully sharply. In the principal–agent situations described previously, for instance, the suggested approach requires that researchers can adequately specify policy determination under the hypothetical extremes of principal and agent full control, that the inputs to these policy response functions and to the function describing monitoring and enforcement costs vary empirically in sample, and it gains the empirical leverage to produce revealing estimates of those parameters only to the degrees that they do so with explanatory power. Then, too, tests of hypotheses regarding the parameters estimated generally tend to weigh that the **x** matter in the way specified against **x** does not matter. The same is true in linear regression or any parametric-modeling approach as well, but linear-interactive models containing *X*, *Z*, and $X \times Z$, for example, will have the linear-additive model nested in them, so empirical evidence could favor that *X* matters linear interactively, linear additively, or not at all. Non-linear models may not always have such intermediate complexity models nested within them. Still, many important substantive problems in comparative politics, and in political science more generally, involve complex, context-conditional relationships, and this approach seems to offer a more theoretically, methodologically, and empirically promising way to address those issues than do alternatives.

The conclusion here can be stated thus: *context matters*, so model it! Before leaving that statement as the terse conclusion of this section, let us again adopt the strategy of writing formally the simplest possible reflection of that broad substantive proposition of complex context-conditionality to explore what is logically possible and what is not with regard to gaining empirical leverage upon it. We start from the most general, broadest interpretation to show why that interpretation of *context matters* thoroughly debars any possibility of learning anything (beyond pure description) from comparative history, by any empirical methodology. From there we work down-

[34] Or, minimally, a *correspondence* (where E(y) has several values for a single given set or sets of values).
[35] See *nl.ado* in Stata™. E-Views™ least squares algorithm, LS, accepts any $f(\mathbf{x}, \boldsymbol{\beta})$ desired.

ward to illuminate what sorts of assumptions, theories, or arguments are necessary or useful in altering that situation, along the way discussing some conjectures about what qualitative empirical analyses do or might do in these regards and mentioning very briefly some statistical procedures of germane utility.

The most general formal expression of the proposition that *context matters* might be:

$$y_{it} = f_{it}(\mathbf{x}_{it}, \boldsymbol{\beta}_{it}, \epsilon_{it});\ \boldsymbol{\epsilon} \sim (0, \Sigma_{it});\ i = 1..N,\ t = 1..T,\ n = NT \tag{11}$$

In this model, an outcome, y, at time t in place i (jointly, in context it) is a function, f_{it}, which is not necessarily linear or additively separable and which may differ across i and/or t, of explanatory factors, $\mathbf{x}_{it}$, which may differ and at least potentially vary across i and/or t, and which relate (not necessarily linear additively or separably; rather, according to f_{it}) to the outcome by parameters, $\boldsymbol{\beta}_{it}$, which may also differ across i and/or t, and of a random component, ϵ, drawn from some probability distribution, not necessarily independently, although with mean zero and some defined variance-covariance across information sets, Σ_{it}, although that (multivariate) distribution could also differ across i and/or t. Note that $\mathbf{x}_{it}$, the explanatory factors, could include temporal and/or spatial lags of any complexity, so, e.g., strategic interdependence and/or path dependence are subsumed as possibilities. Thus, (11) interprets *context matters* fully generally and broadly. From this formulation, we see that this broadest interpretation gives K (the number of β's) plus $\frac{1}{2}(NT)^2 + \frac{1}{2}NT$ (the number of unique parameters in each variance-covariance matrix for the random component) total parameters *per function* to learn ***from each information set observed***. If this is our understanding of *context matters*, then we simply cannot learn anything from comparative history because each piece of information observed empirically would come with many, many times that unit of information to learn. Social scientists therefore must reduce this parameterization, i.e., impose tighter structure on this formulation.[36] The imposed additional structure would ideally come from theory and/or substance or, failing that, practicality, but in any event some assumptions must be made, regardless of empirical methodology, no matter how many contexts we analyze or how closely we analyze them.

To begin, we usually assume $f_{it}(\cdot) = f(\cdot)\ \forall i,t$, i.e., the same function relates $\boldsymbol{X}_{it,it}$, and ϵ_{it} to y in all contexts, opting instead to parameterize variations in the effects of variables across contexts as we described doing previously in this section. Note that this subsumes (i) allowing the effects to vary across but not within certain groups of contexts (e.g., across countries but not over time, or vice versa) and (ii) allowing the set of relevant factors to vary across contexts. Likewise, we always assume $\Sigma_{it} = \Sigma\ \forall i,t$, i.e., that the random component of the outcome in each context is a draw from a distribution with some variance-covariance across contexts, but this variance-covariance across contexts is itself fixed across contexts. Either

[36] Indeed, even though historians might say that those who do not learn from history are doomed to repeat it, they would simply be wrong about that if they also believed "each context is unique" to this full extent because, then, any historical episode would, as this formulation shows, be entirely *sui generis*.

assumption may be relaxed somewhat if other restrictions open sufficient degrees of freedom to do so. This still leaves $K(NT) + 1/2(^{\mathrm{nt}})^2 + 1/2^{\mathrm{nt}}$ per NT contexts, or $K+1/2(^{\mathrm{nt}} + 1)$ per context, which is still way, way too many to learn anything from comparative history. We must, of course, get to less than NT things to learn per NT contexts before we can learn anything (beyond description of a now-irrelevant past) from empirical observation of any kind.

Next, we can assume the $\boldsymbol{\beta}$ constant across contexts. This still leaves $K/(\mathrm{NT})+1/2(\mathrm{NT}+1)$ per context, which is still much larger than one; however, this also may be stronger than needed and definitely is stronger than desired given the centrality of context-conditionality to comparative politics. We can, in fact, allow $\boldsymbol{\beta}_{it} = g(\mathbf{z}, \boldsymbol{\gamma}, \eta_{it})$, that is, we can model the contextual variation in the effects of variables as a function, g, of other variables, $\mathbf{z}$, with parameters, $\boldsymbol{\gamma}$, which is what we discussed above. We could even allow a random component, η_{it}, i.e., stochastic variation, in the effects of variables, which produces the random effects or random coefficients model that we will mention again below (see, e.g., *Political Analysis* 2005: 13 (4)). We can do these things, though, only provided the parameters to learn in g, including those in the variance-covariance of the random effect component(s) if any, remain less than the number of contexts, NT, minus K, the number of $\boldsymbol{\beta}$, minus the number of parameters involved in specifying Σ^{γ} and Σ^{ϵ}.

Alone or together, these steps will not suffice, however, because even a single variance-covariance of the stochastic components across the NT contexts, Σ^{ϵ}, by itself contains up to $1/2(^{\mathrm{nt}})^2 + 1/2\mathrm{nt}$ unique parameters, which far exceeds NT. (That's $1/2(\mathrm{NT}+1)$ per context!) The structure of this variance-covariance gives our explicit statement as to how each context is related to and informed by all other contexts at different times and/or places. In an annual, cross-national information set, e.g., if how France 1972 relates to France 1971 may differ arbitrarily from how 1971 relates to 1970, and how France 1986 relates to Germany 1986 from how France 1987 relates to Germany 1987, etc., then degrees of freedom are negative, and comparative history can offer nothing beyond description of exclusively photographic interest. One must at this point emphasize once again that the challenge, and the inability of gathering more empirical information or of considering some contexts more closely to redress it, and so the necessity of restrictive assumptions, do not arise because we have written the proposition down formally, and do not go away if we do not. Thus, for example, when a scholar claims to have learned something useful for contexts beyond those now past and irrelevant episodes in some one or few closely studied context(s), s/he may be correct, but, if so, that must be because some assumptions restraining the generality of the allowed variance-covariance across contexts have been imposed (at least implicitly). Empirical inference was not logically possible otherwise. Perhaps seeing explicitly some of the common restrictions in standard statistical analyses would be helpful here. We begin by describing the contents of the variance-covariance matrix, Σ^{ϵ}, for a generic time series cross-section of information (a structure befitting comparative politics):[37]

[37] For tractability, (12) shows just a small, 2-unit-T-time-period example; more generally, the block structure of the variance-covariance would expand horizontally and vertically for N units.

$$
\begin{aligned}
\mathbf{V}(\boldsymbol{\epsilon}) \equiv \mathbf{V}(\mathbf{y}|\mathbf{X}) \equiv \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega} \\
= \sigma^2 \times \begin{bmatrix}
\omega_{1,1}^2 & \omega_{1,12} & \omega_{1,13} & \cdots & \omega_{1,1T} & \omega_{12,11} & \omega_{12,12} & \omega_{12,13} & \cdots & \omega_{12,1T} \\
\omega_{1,21} & \omega_{1,2}^2 & & & \vdots & \omega_{12,21} & \omega_{12,22} & & & \vdots \\
\omega_{1,31} & & \omega_{1,3}^2 & & \vdots & \omega_{12,31} & & \omega_{12,33} & & \vdots \\
\vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots \\
\omega_{1,T1} & \cdots & \cdots & \cdots & \omega_{1,T}^2 & \omega_{12,T1} & \cdots & \cdots & \cdots & \omega_{12,TT} \\
\omega_{21,11} & \omega_{21,12} & \omega_{21,13} & \cdots & \omega_{21,1T} & \omega_{2,1}^2 & \omega_{2,12} & \omega_{2,13} & \cdots & \omega_{2,1T} \\
\omega_{21,21} & \omega_{21,22} & & & \vdots & \omega_{2,21} & \omega_{2,2}^2 & & & \vdots \\
\omega_{21,31} & & \omega_{21,33} & & \vdots & \omega_{2,31} & & \omega_{2,3}^2 & & \vdots \\
\vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots \\
\omega_{21,T1} & \cdots & \cdots & \cdots & \omega_{21,TT} & \omega_{2,T1} & \cdots & \cdots & \cdots & \omega_{2,T}^2
\end{bmatrix}
\end{aligned}
\tag{12}
$$

Let us describe the elements of this matrix to understand better that about which we must make assumptions. We first underscore that this is the variance-covariance matrix of the stochastic component, i.e., of the residual after netting our model of the systematic component. If one theoretically or substantively expects some covariation of observations, the first move usually should be to try to model this *systematic* expectation in the *systematic* component. Given that, note next that the two blocks on the prime diagonal (top left to bottom right) are the variance-covariance matrices for units one and two respectively. The prime-diagonal elements of those submatrices give the relative variances of information sets *1* to *T* within that unit. The off-diagonal elements of these prime-diagonal blocks give the covariances of the corresponding time period observations; e.g., $\omega_{2,4}$ is the covariance of the second-period with the fourth-period observation. The entire matrix and each block submatrix are symmetric (mirrored above and below their prime diagonals) because, e.g., $\omega_{2,4} \equiv \omega_{4,2}$. The off-diagonal blocks give cross-unit covariances. The prime diagonals of these blocks give the contemporaneous (same time period) covariances, and their off diagonals give the cross-temporal cross-unit covariances. For example, $\omega_{21,11} = \omega_{11,21}$ is contemporaneous covariance in the first period across these two units, and $\omega_{21,13} = \omega_{13,21}$ is the covariance of the second unit's first period with the first unit's third period. If we leave each of these $1/2NT(NT+1)$ elements to differ arbitrarily, then, to put it in simple language: no time and place could offer empirical information relevant to any other time and place.

The most stringent of the common assumptions to redress this is *sphericity*, i.e., that we have sufficiently modeled in the systematic component all sources of covariation and non-constant variance across contexts, which is what ordinary least squares regression assumes: $\sigma^2\boldsymbol{\Omega} = \sigma^2\mathbf{I}$. This reduces the parameterization from $1/2\,NT(NT+1)$ terms to just one term to learn per NT pieces of information and thereby earns us $NT-1$ degrees of freedom to spend on enriching the model of the systematic

component with interesting multicausality and context-conditionality.[38] However, such a stark assumption—it amounts to a claim that, net of our systematic-component model, each observation is an independent draw from one constant distribution[39]—may often seem implausible. Another common assumption is *panel heteroskedasticity*, which holds that sources of covariation are sufficiently modeled, but variances of the distributions from which observations are (conditional-) independently drawn may differ: $V(\epsilon_{it}) = \sigma_i^2$. This eats N degrees of freedom, leaving $NT - N$ for enriching the systematic-component model. In the temporal dimension, a common parameterization is that stochastic components correlate directly from one observation to the next, giving $\epsilon_{it} = \rho\epsilon_{i,t-1} + \upsilon$ and two parameters if this correlation, ρ, is assumed the same across units or $\epsilon_{it} = \rho_i\epsilon_{i,t-1} + \upsilon$ and $N+1$ parameters if it is assumed to differ arbitrarily, leaving $NT-2$ or NT−N−1 degrees of freedom. The Parks–Kmenta procedure that Beck and Katz (1995) so influentially exiled from comparative politics practice added contemporaneous correlation of the form $\sigma_{ij} = \sigma_{ji}$, one such for each dyad ij, to unit-specific ρ_i and σ_i^2. This gives $2N+1/2\ N(N-1)$ parameters, meaning that positive degrees of freedom require $T>N$. To obtain reasonable estimates of parameter estimate uncertainty, $T \gg 2N$ seems necessary.[40]

Many other plausible parameterizations are imaginable and feasible. Researchers can in fact structure Σ however they like—that is, they can assume about the relative covariances and variances of their observed information whatever follows theoretically substantively—provided (i) that these assumptions, combined with stances taken on the other elements of their model as discussed above, yield fewer parameters to learn than pieces of information available and (ii) that researchers understand and accept that increased uncertainty and greater demand for more numerous and usefully variant information sets is unavoidably concomitant with allowing greater complexity. As always, writing the model formally clarifies, but does not create, the logical requirements for empirical evaluation just noted. A question, neither rhetorical nor intended as implicit jibe, then, is what sorts of assumptions are qualitative analysts offering about, e.g., the amount of novel information contained in each of temporally adjacent episodes from with a single context? The answers are not always clear, yet such assumptions logically must be being held whenever inferences about the empirical validity or utility of theoretical propositions are being drawn, by whatever methodological approach, from what is observed. Methodological research seeking to close qualitative-quantitative divides needs to provide better answers to such questions.

Working from the most general model downward, as we just did, counters a parsimony favoring method of theoretical development and empirical evaluation

[38] As our parents would say, however, don't spend it all on enriching the substantive component because we will need a large number to spare to endow our estimates with any appreciable certainty!

[39] More exactly, independently from one distribution or from distributions all having the same variance.

[40] The latter is because Parks–Kmenta, like any standard FGLS procedure, ignores parameters estimated in $\hat{\Sigma}$.

that would work from simplest upward. As guides to empirical practice, the latter requires evidence to sustain increasing model complexity; the former requires evidence to support reducing it. Franzese (2005, 2006) offers discussions of the possible in general empirical evaluation of positive theory in political science working from simplest models upward that are similar in motivation and spirit to the present discussion, but more technical. Conclusions therein can be summarized thus. If one can model the theoretical/substantive reasons for deviation from the classical linear regression model (CLRM), do so, and, if and insofar as one does so successfully, the strategy is optimal in all regards. Insofar as possible, "Model It!" thusly in the first-moment, $E(\mathbf{y})$, i.e., the systematic component, for two reasons: (i) most usually, theoretical/substantive information regards the systematic, not the stochastic, components, and (ii) observationally, the only information one may obtain on the stochastic component (i.e., second moment) is conditional on information from the first moment, i.e., on one's model of the systematic component. So insufficiencies in the first-moment model will seem incorrectly to be violations of second-moment CLRM assumptions, and creditworthiness of empirical results regarding second-moment relationships is governed by that of the first-moment model. However, some theoretical propositions and substantive information do directly regard second moments (variances and covariances); e.g., education or lack thereof may induce survey respondents to answer more variably (second moment) rather than, or as well as, differently (first moment). In such cases, "Model It!" is still the advice; in second moment, this means to model a reduced parameterization of $\boldsymbol{\Omega}$. Finally, insofar as one fails to model sufficiently the theoretically/substantively expected deviations from the CLRM, problems arise essentially as omitted variable bias in worst cases, but as "just" inefficiency and incorrect standard errors in many other cases. Redresses, mostly partial and/or imperfect, of deficiencies in implementation of the "Model It!" strategy include sandwich estimators of the variance-covariance of parameter estimates (e.g., so-called *robust* standard errors, such as Beck and Katz's famous PCSE), FGLS (e.g., Parks–Kmenta), fixed-effects estimators (aka dummy variable regression), shrinkage estimators (like random effects or random coefficients, aka hierarchical, models).

The upshot of this section is that complex context-conditionality not only does not debar quantitative approaches to *general* empirical evaluation, but rather tends to demand them because empirical evaluation of theories that nest complex context-conditionality within general models that explain the nature of the conditionality across contexts logically necessarily places greater demands on precise calculation of multiple partial associations and requires greater variation across more numerous sets of information. The advantages of closer scrutiny of fewer contexts likely lie more in exploration and (potential) discovery of such conditionality, a process closer to theoretical development and refinement and, empirically, to sensitivity analysis than to general empirical evaluation.

5. Endogeneity: Almost Everything Causes Almost Everything Else

The third, final, and in many ways largest of the fundamental challenges for empirical evaluation of positive theory in social science, the third way in which *context matters*, is the ubiquitous issue of endogeneity (aka, simultaneity, selection, bi- or multidirectional causality);[41] i.e., more often than not, X's cause Y's and Y's cause X's: $X \Leftrightarrow Y$. Section 2, for instance, used an example of the relationship of economic development to democratic quality, being careful to speak of association rather than cause and effect because correlation between the two could as easily derive from a causal relationship from development to democracy, from democracy to development, or both (not to mention the spurious possibilities). The likelihood of coups, social instability, and regime change, too, may increase in economic inequality or hardship, but sociopolitical instability, or expectation thereof, also hinders investment and so development (e.g., Przeworski et al. 2000; Londregan and Poole 1990). Theorized vicious or virtuous circles are another class of examples common in comparative politics. For example, trust or social capital begets well-functioning sociopolitical institutions, and such effective functionality renders rational popular expectations of cooperation from others and the public, i.e., social capital (e.g., Putnam 1993). In a relatively neglected political-economy example, the temporal proximity of an election may induce incumbents to manipulate policies in various ways to cultivate current favor from informationally disadvantaged publics, yet public support (or trends expected therein) can strongly affect the timing of elections in some parliamentary systems (e.g., Smith 2004). Further examples are easy to amass since most everything causes most everything else in society, polity, and economy, but suffice it to note how often one study's dependent and independent variables reverse roles in another (or later in the same study, and sometimes even simultaneously), even if direct references to bi- or multidirectional causality are not made.

We adopt once more the strategy of formalizing the positive propositions inherent in core tenets of comparative politics as a means of understanding better possible empirical leverage on them and alternative empirical-methodological approaches thereto. Let us start with a simple, two variable system: $X \Leftrightarrow Y$. Suppose a researcher only knows or is willing to argue that democracy, X, spurs development, Y, and development favors democracy; simplifying (inconsequentially for the conclusions) that both relations are linear, this gives:

$$\begin{aligned} & Y = a \times X;\ X = b \times Y \\ & \Rightarrow Y = a \times (b \times Y) \\ & \Rightarrow ab = 1 \text{ or } a = b^{-1} \end{aligned} \tag{13}$$

[41] NB, sets of definitions and usages of these terms are not entirely consistent across texts, even just within the formal statistics literature. One commonly used and useful set defines endogeneity as any covariance-of-regressor-with-residual problem, of which simultaneity and its exact synonym bi-/multidirectional causality, and selection, which is very nearly synonymous as well, are subsets. All four are used synonymously here.

If the first line were all we could say about this system, i.e., that Y is a function of (depends on) X and vice versa, then we could not say much or learn much from comparative history about this simultaneous relationship because its parameters, a and b, are *unidentified*. The last line of (13) reveals this; absent some further information, *any* a and b such that $a = b^{-1}$ is equally consistent with the circular proposition $X \Leftrightarrow Y$. An infinite set of (a,b) satisfies this relation, although not just any a and b. Valid (a,b) must satisfy $a = b^{-1}$; the specifications in the first line of (13) plus $X \Leftrightarrow Y$ get us this far, at least. However, without more theory, we cannot get any further toward establishing the (causal) *effect* of X on Y or of Y on X, i.e., toward gauging a or b, no matter how many sets of information we might obtain or how closely we might explore any given set(s) of information. We (logically) necessarily require *extra-empirical* information (from the usual candidates: theory, substance, or assumption).

We can see the situation graphically if we consider, e.g., a scatterplot of "democracy" (Freedom House political plus civil liberties) and "development" (real GDP per capita). In Figure 2.2, the long, downward-sloping line gives the simple linear association (regression) of democracy with (on) development. This association, however, reflects neither the effect of development on democracy, $Dem = f(Dev)$, nor the effect of democracy on development, $Dev = g(Dem)$, but rather the particular mix of those two causal relationships that happen to have manifested across this set of contexts (sample). That is, each data point (context) gives the intersection of the two lines that happens to have occurred in that country in 1980—Figure 2.2 shows a hypothetical pair of these lines crossing through Uganda 1980—and both lines likely shift, for many reasons (multicausality and context-conditionality), from one datum to the next. As drawn in Figure 2.2, the inferential error made by interpreting the

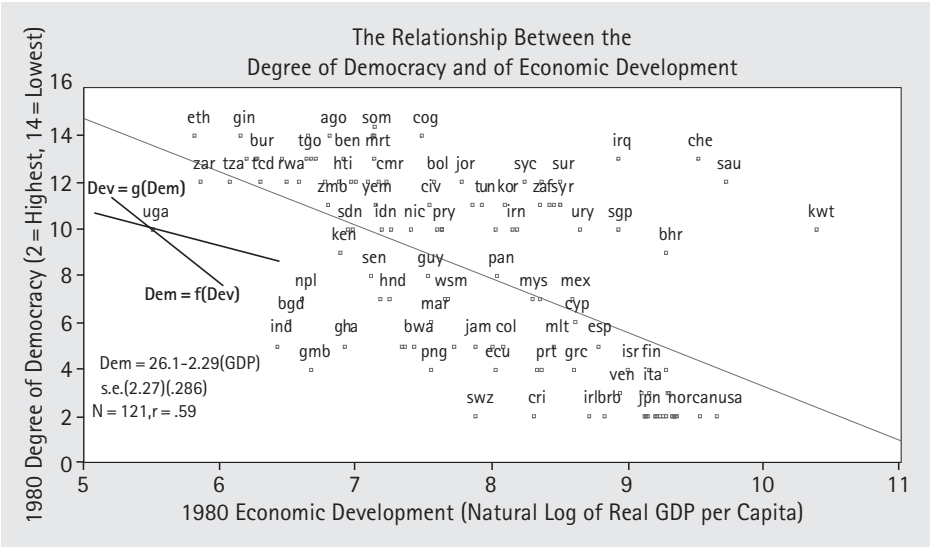

Fig. 2.2 Empirical vs. Causal Relationships of Democracy and Development

observed association of democracy with development as the effect of development on democracy (i.e., the simultaneity bias) is appreciable. As drawn, one overestimates the impact of development on democracy, which is rather small in this example, by neglecting the rather sizeable (in this example) effect from democracy on development. The bias could be much larger still if disparity between the slopes of the multiple causal relationships were greater (going in opposite directions, for example, such as in situations paralleling supply and demand as a function of price). Biases from interpreting observed association as causal effect is not a function of how one observes the association, nor does it depend on how closely or often one observes it. Accordingly, all of the strategies that social (and other) scientists have brought to bear upon identifying causality—(i) from imposing theoretically derived structure (assumptions) on the empirical estimation model, (ii) to experimentation in lab, survey, or field, (iii) to (propensity score) matching, (iv) to vector autoregression, to (v) "process tracing" or "causal process observations" and the like—ultimately work by (and succeed only insofar as) adding/imposing (correct) *extra-empirical* information on the empirical analysis.[42] Furthermore, the correctness of the logically necessary *extra-empirical* information cannot be empirically tested directly. Przeworski (this volume) discusses causal analysis in general and most of these five broad classes of strategies for identifying causal effects in particular sufficiently for most of our present purposes, so we focus here on issues receiving less attention there. Let us, then, characterize more formally the general challenge of ubiquitous endogeneity—most everything causing most everything else—to clarify how much extra information is needed and in what forms it can come, briefly discussing these five broad classes of strategies for identifying causal effects along the way.

Consider a system of M endogenous outcomes, $\mathbf{y}$, that depend—linear additively and separably for the sake of simplicity—on each other and on K exogenous explanatory factors, $\mathbf{x}$, plus a stochastic error, $\boldsymbol{\epsilon}$, for each. In matrix notation, we can write such a system of M equations *for one empirical observation of a context*, i, as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_m \end{bmatrix}_i' \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm} \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ x3 \\ \vdots \\ x_k \end{bmatrix}_i' \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{k1} & \beta_{k2} & \ldots & \beta_{km} \end{bmatrix} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \vdots \\ \epsilon_m \end{bmatrix}_i$$

In matrix notation, the system may be written compactly as:

$$\mathbf{y}_i' \Gamma + \mathbf{x}_i' \mathbf{B} = \boldsymbol{\epsilon}_i \tag{14}$$

In (14), Γ is an MxM matrix of up to M^2 coefficients on each y in each y's equation; $\mathbf{B}$ is an $K \times M$ matrix of each of the K exogenous variables' coefficients in each of the M equations; and $\boldsymbol{\epsilon}$ is an $M \times 1$ vector of stochastic components of the endogenous

[42] Whether the extra-empirical information is added in Bayesian or Classical manner affects our discussion little, so we can safely side step that involved discussion.

variables, with a corresponding $M \times M$ matrix of variances and covariances across the equations. Variance-covariance matrices are symmetric, so $V(\epsilon) \equiv \Sigma$ has $1/2\, M^2 + 1/2\, M$ unique elements. So, with each information set an empirical researcher observes on a set of M endogenous variables, she has, at most general, up to $M^2 + KM + 1/2M(M+1)$ parameters to learn. Luckily, the associations observed between the variables in that one information set, while not the causal effects sought (rather, some complicated mix of them as, e.g., Figure 2.2 illustrated in a simple, two-variable case), will nonetheless reduce the amount of extra information she needs to bring to the system to "tie it down" (i.e., identify it). In particular, if she simply regressed the M y's on the K x's, she would find some $\mathbf{y} = \mathbf{x}'\boldsymbol{\Pi} + \boldsymbol{\nu}$, with $V(\boldsymbol{\nu}) = \Theta$, which provides $KxM+1/2M(M+1)$ pieces of information—namely, observed associations, variances, and covariances across equations, which are useful information, but not the causal effects sought. So, in general, for a system of M endogenous variables, we need M^2 additional, extra-empirical pieces of information to get valid estimates, qualitative or quantitative, of effects. The first M of those are essentially automatic. The diagonal elements in Γ are the coefficients on each y in its own equation, and so amount to arbitrary scaling coefficients, which we always set to one (because explaining y is more direct than explaining, e.g., $4y$).

Practically, then, a system of M equations requires $M(M-1)$ extra-empirical pieces of information, i.e., theoretical or substantive (and empirically untestable) restrictions.[43] For example, our system of two endogenous variables, democracy and development, requires two extra-empirical pieces of information. In what forms can these come? *Identities:* If some things are true for certain, say as accounting or other identities, then one can impose these facts rather than attempt to estimate them. *Exclusions:* factors from among the set of $\mathbf{x}$ or $\mathbf{y}$ that are excludable (by assumption) from some equation—identically, whose coefficient in those equations can be fixed to zero. Such "instrumental variables" assumptions are the most common econometric strategy. *Coefficient restrictions:* Exclusions can be generalized to any sort of parametric restrictions on possible coefficients for $\mathbf{x}$ or $\mathbf{y}$ variables, such as that two or more coefficients must be equal or proportional across equations. *Functional form:* the intuition is complicated for how functional-form information can help identify systems and how much applications of such strategies can buy empirical researchers. In essence, if one is willing to impose that some variable matters according to a specifically shaped function in one equation and (a) differently shaped one(s) in others, then one can leverage the difference between the imposed shapes of these relationships to help identify the system (instead of or in addition to other restrictions). *Stochastic-component variance-covariance restrictions:* How this sort of additional information, which, in different ways, is what experimentation and matching, on the one hand, and vector autoregression on the other primarily employ, works is also complicated. Experimentation and matching methods randomize on unobservables as Przeworski (this volume) describes. If such match or randomization is achieved, which, as he also notes, is not directly testable, this amounts to restricting the

[43] This is just the rank condition; it ensures that enough information has been added to identify M equations. The more complicated order condition ensures that *each* equation is identified. Jointly, the rank and order conditions are necessary and sufficient for identification.

variance-covariance matrix to the single parameter of the variance of that *independent* random component. Experiments also manipulate potential causes, ensuring safe exclusion of treatments applied in one equation from other system equations. The general rule for what imposition of the extra-empirical pieces of information buys empirical researchers is that each immovable fact or each fact rendered immovable by assumption yields one parameter identified.

Graphical intuition for how exclusions (and other coefficient restrictions) work may be seen by imagining in Figure 2.2 that one had another variable, say climate, that affected, say, economic development but that did not affect democracy except insofar as it affected development. Such an explanatory factor would enter the equation for development, but not directly that for democracy. As such, variation across contexts in climate would shift the $Dev = g(Dem)$ function around, but would not shift the $Dem = f(Dev)$ functions, thereby tracing out for the researcher that $Dem = f(Dev)$ causal function. Conversely, if something could be found that in parallel manner entered the $Dem = f(Dev)$ but not the $Dev = g(Dem)$ function, then the $Dev = g(Dem)$ causal function could be traced. The two extra-empirical pieces of information required to identify both equations of the system would be these untestable exclusion assumptions. Practically, in brief, one would regress development on climate and then use that prediction rather than development itself as a regressor in the democracy equation (this is the instrumental variables by two-stage least squares method).

Vector autoregression and related techniques, for their part, amount to sophisticated applications of what might be called the "poor man's exogeneity:" history. Namely, things that happened in the past are assumed exogenous to what happened later. In the strictest sense, this must be true. However, in social science practice, expectations of the future can cause outcomes today, and, if the empirical model does not capture these expectations sufficiently, then future values of outcomes can seem to cause present ones. Similarly, expectations of contemporaneous outcomes or actions of others can shape one's own actions contemporaneously, so time lag is not always necessary for cause to induce effect in social science. Contemporaneous response can occur. Similarly, if the empirical models insufficiently capture temporal dynamics, then those inadequacies can leave future observations conditionally correlated with present ones, and so induce endogeneity. Many applications of instrumental variables strategies employ this poor man's exogeneity (i.e., endogenous variables are time-lagged and declared exogenous), as do many qualitative strategies, one suspects. The sophisticated way in which vector autoregression uses time, though, to describe its practice briefly, is to regress each endogenous variable on some number of its own time lags and of all other endogenous variables. The residuals from these regressions are then, by construction, (linearly) inexplicable by lags of any of the endogenous variables. One could in principle then use these dynamic models as estimated to trace responses (called *impulse-response functions*) of all the endogenous variables to these "inexplicable shocks" (called *innovations*). The remaining problem, though, is to determine to which variables to allocate the covariation across the equations' residuals/innovations. VAR resolves this by temporally ordering the responses, i.e., positing that some variables adjust more quickly than others.

This restricts the temporal feedback between the jointly endogenous variables and thereby identifies, not the system directly, but these impulse responses and related estimation outputs such as Granger-causality test statistics or explained-variance decompositions.

Przeworski describes how experimentation and matching achieve randomization and the issues involved therein. One might only add emphasis on two aspects of his discussion for our purposes. First, randomization requires large samples to achieve its beneficial effects reliably. That is, even if, in making causal comparisons, we believe that unobserved factors are drawn independent-randomly, having just one or very few such comparisons would render the estimate's unbiasedness (being correct on average or in expectation *across many draws*) cold comfort and would not suffice to draw any solace from the consistency (being exactly right with no uncertainty *as the number of draws approaches infinity*). Thus, "quasi-experimental" and matching-like logics are problematic bases for drawing causal inference from small numbers of information sets. Second, the stable unit treatment value assumption (SUTVA) underlying matching methods, in particular (but not solely) the implication of SUTVA that one unit's receipt of the treatment does not affect the receipt by, or the value of the treatment to, other units, seems implausible for many comparative politics applications. One could hardly imagine, for instance, that the nature of the regime in one country had no effect on regimes or their effects in others, as matching methods would require for valid estimates from an observational study of the effects of regime type.

Tracing episodes through "causal-process observations" (e.g., Brady and Collier 2004; Bennett and Elman 2006) or similar methods of close and careful qualitative analysis (e.g., Hall 2003) have been advocated as particularly effective strategies for assessing (complex) causality. All of the potential sources of additional information logically necessary to evaluate and gauge causality are as available to qualitative as to quantitative empirical methods, so this may be. Indeed, since the necessary information is *extra-empirical*, whether one employs qualitative or quantitative approaches while imposing that extra information is largely irrelevant; conversely, though, the choice of qualitative or quantitative approach will not *ipso facto* provide the logically necessary extra-empirical information. Simply tracing some process (set of episodes) through time to establish which movements or episodes occur in what order, for example, would seem to be familiar assertion of the poor man's exogeneity, and so to come with that instrument's drawbacks or caveats regarding expectations and the need to specify very precisely and sufficiently the dynamics of the process. Process tracing may also involve analogs to experimental or matching analyses if, for example, closer scrutiny enables the researcher to hold with greater substantive-theoretical certainty that particular moving factors in their account could only have moved exogenously.[44] However, notable relative weaknesses plague qualitative analogs

[44] Empirical certainty, on the other hand, that some factors have moved exogenously should probably be seen as problematic if not impossible to ascertain because any observed associations, seen closely or distantly, in numerous or in scant contexts, can be misleading about causality and so endogeneity and exogeneity.

to experimental and instrumental strategies: namely, the precision in model specification necessary to effective instrumental strategies is not a relative strength of the approach, and the randomization that undergirds experimental and matching approaches provides only very weak basis for confidence in comparisons of few contexts. More promisingly analogous, therefore, may be the vector autoregression approach of distilling innovations from that which is predictable from raw dynamics of the endogenous variables, imposing a temporal ordering to the incidence of those impulses, and tracing responses thereto. If so, considerable work translating the logic of that identification strategy to something understandable in qualitative analysis terms remains. Finally, note that the issues discussed in previous sections regarding quality–quantity tradeoffs, multicausality and the difficulty of ascertaining partial associations, and the challenge of modeling and assessing (complex) context-conditionality—all pervade and compound this already thorniest of challenges for empirical evaluation of social science theory, ubiquitous (and perhaps complex) endogeneity. To evaluate causality and gauge causal effects from comparative history, in other words, requires effective redress of all these challenges. Once again, then, vis-à-vis general empirical evaluation, writing explicitly the logical challenges for empirical evaluation associated with this central tenet of *context matters*—most everything causes most everything else—seems to indicate that, far from debarring quantitative analysis, ubiquitous and complex endogeneity tends to demand it.

6. Conclusion: Context Matters, so Model It!

In an influential critique of empirical practice in comparative politics that similarly emphasizes that *context matters*, Hall (2003) argues for close empirical analysis of several (i.e., more than the one or very few of one end of current practice, less than the many of the other), raising these concerns about regression analysis in comparative politics:

> . . . [1] the types of regression analyses commonly used to study comparative politics provide valid support for causal inferences only if the causal relations they are examining meet a rigorous set of assumptions (see Wallerstein 2000). [2] In general, this method assumes unit homogeneity, which is to say that, other things being equal, a change in the value of a causal variable x will produce a corresponding change in the value of the outcome variable y of the same magnitude across all the cases. [3] It assumes no systematic correlation between the causal variables included in the analysis and other causal variables omitted from it. [4] It assumes that all the relevant interaction effects among the causal variables have been captured by interaction terms in the regression. [5] It assumes that the cases are fully independent, such that [6] the values of the causal variables in one case are unaffected by the value of the causal variables or outcomes in other cases. [7] Although instrumental variables can sometimes be used, most regression analyses assume that there is no reciprocal causation, i.e. that the causal variables are unaffected by the dependent variable . . .

Here we have engaged and continued the discussion by expressing formally the specific multicausal, context-conditional, and ubiquitous-endogeneity propositions entailed in the *context matters* central tenet of comparative politics. We have done so to clarify the logical requirements of empirical evaluation from comparative history in general, and some of the specific approaches to fulfilling those requirements that quantitative methods employ. That, in turn, may help clarify what the corresponding moves to fulfill those requirements might be in qualitative methods and help to define and to characterize more sharply the terms of any tradeoffs between the approaches. Thus, we have seen that, regarding Hall's concerns: [1] rigorous assumptions are necessary to valid support for causal inferences by any methodology; moreover, any set of assumptions chosen must achieve the same things in terms of parameter reduction and the like to allow meaningful empirical inference from comparative history. [2] Context-conditionality, regardless of complexity, can be modeled, estimated, and interpreted quantitatively provided the context-conditional propositions are logically consistent and that sufficient empirical information logically could exist to gauge them;[45] if insufficient information actually exists in comparative history to estimate these relations well, it is unlikely that restriction to narrower sets of contexts will add the needed further contextual variation, and increased quality of information is unlikely to compensate sufficiently. [3], [4] If potential causal factors, whether interactive or simple causes, are omitted from analysis, whether qualitative or quantitative, inferences will be biased if the omitted are indeed causal and also correlate with the included. Likewise, however, equally valid concerns should be considered regarding excess complication of empirical analyses (Achen 2002). [5], [6] One need not assume independence of outcomes[46] across contexts (e.g., time and/or place), but, by any method of analysis, one must assume some pattern of correlation across contexts that reduces the information needed to gauge and account for that interdependence of outcomes sufficiently to leave enough free information from the available comparative history to infer something also about the systematic aspect of the proposition being analyzed empirically.[47] [7] The severe challenge that reciprocal causation raises for empirical evaluation requires *extra-empirical* information to

[45] Later, Hall lists some manifestations of contextual complexity seen as challenges for regression analysis in particular. "*i. We find instances in which an increase in x (level of economic development) causes an increase in y (movement toward democracy) in some cases but does not have this effect in others, where y is caused by an entirely different set of variables, w. ii. We find cases in which an increase in x (social democratic governance) is associated with an increase in y (social spending) at one point in time, t1, but not at another point in time, t2. iii. We find instances in which an increase in x (social protest) causes an outcome y (government turnover) in some cases but an entirely different outcome (repression) in other cases. iv. We find instances in which an outcome y (successful wage coordination) depends on the value of many other variables, v (union density) w (social democratic governance), and x (social policy regime), whose values are in turn jointly dependent on each other. v. We find cases in which increases in x (support for democracy) increase y (the stability of democracy) and in which increases in y also tend to increase x.*" Each of these, e.g., is easily written as an estimable empirical model. The open-endedness of some qualitative analysis may allow researchers to discover signs of such context-conditionality, but whatever systematically context-conditional propositions may emerge cannot be well evaluated empirically in the same discovery process.

[46] The dependence or independence of explanatory factors is not an issue (for any method of analysis) unless these explanatory factors are also endogenous to outcomes.

resolve; therefore, no particular empirical methodology or approach brings such information by itself or enjoys any inherent advantage in generating it. Using such extra-empirical information effectively to explore endogenous relationships, however, does seem to require a certain mathematical precision in processing the empirical and extra-empirical information that may favor quantitative strategies. The same seems true regarding the partialing of evidence related to multicausal and/or context-conditional relationships in general.

The biggest and fundamental challenges for empirical evaluation in comparative politics—multicausality, context-conditionality, endogeneity—inhere logically in the nature of the theoretical processes argued to be present and being considered for empirical evaluation: *context matters*. That implies that any approach we may offer for obtaining useful empirical leverage on such propositions must somehow address these same logically inherent challenges. By any approach, if we believe we have learned something useful as anything more than a photograph of some specific scenario(s), a photo that is wholly useless in any other scenario (tomorrow in the same exact geographic, cultural, strategic, etc., context, for example), then we must have offered, implicitly or explicitly, some redress of those challenges—and always and everywhere, it will be partial redress. In other words, if one claims to have learned something from the comparative historical record that is of use for anything beyond solely describing that now-gone situation—and explaining that situation is as beyond describing it as is understanding by that analysis of that situation something useful in other, related situations—and regardless of whether one has used that comparative historical record in statistical or some other kind of analysis to get this *understanding-beyond-description*—then one must assume, implicitly or explicitly, something about how these multicausal, context-conditional, and/or endogenous relations in this scenario relate to those in other scenarios. (Indeed, even photographic description may be impossible without some minimal stands on these issues.)

One can tell what these necessary parameter-reducing assumptions are in a given statistical model—e.g., that the effect of X is a constant in all contexts like those in the sample, or that the effect of X depends on (only) Z (linear additively), etc. These sorts of necessary assumptions tend to be similar but more flatteringly put: *flexible*; pejoratively put: *arbitrary*; perhaps fairly put: *subjective*—in qualitative methods. In any event, this sort of flexibility is not a virtue in providing general empirical evaluation. Moreover, we have increasingly found over the course of our discussion here that contextual complexity, far from arguing for closer analysis of narrower sets of contexts, tends almost universally to argue strongly against it *for purposes of general empirical evaluation*. We hasten here to reiterate, as we had at the start, that qualitative analysis is an essential part of the scientific enterprise. These methods have particular advantages that seem, however, to have little to do with general empirical evaluation of multicausal, context-conditional, and ubiquitously endogenous relationships. Their great advantages seem

[47] Moreover, assumptions of identical data-generating processes or independence in statistical analysis, if the analyst does make such assumptions, regard Y|X and not Y. That is, these assumptions apply to the outcome and scenarios being compared *controlling for* the actual empirical model on offer. If one thinks context alters the effects of some X, for example, then one can and should model this modification of effects in Y|X.

instead to lie most heavily in ascertaining and validating conceptualization and measurement quality, in exploring applicability, sensitivity, and robustness, and in developing and refining theory. The advantages lie in those equally essential parts of the iterative continuum of theoretical development and empirical analysis rather than in general empirical evaluation—after all, narrower deeper focus is precisely *not* broad and general. This theory-building/empirical-evaluation iteration is also more continuous in qualitative and more discrete in quantitative analyses, which greater merging of the acts of theory building and empirical evaluation in the former is also not a virtue in terms of general empirical evaluation. Conversely, the weaker points tend to lie precisely in the areas of empirical evaluation of those theories given those given concepts and measures, in fact especially in empirical evaluation of multicausal, context-conditional, endogenous relations. Put more crudely than perhaps it should be: qualitative empirical analyses tend to be robustness checks, sensitivity analyses, stress tests, and field tests—after having built a new power tool and tested it in the lab to show its general safety and efficacy, one also gives it to some carpenters to use in the field to discover whether it is ultimately useful!—more than general tests. Far greater strengths for the approach lie in its potential for theory building and refinement.

Viewing this comparison of the relative effectiveness of the approaches in different aspects of the broader scientific endeavor as a competition is rather pointless, though. Provided that we all share the same or very similar overarching goals—"theoretically and empirically useful understandings" may perhaps be reasonably uncontroversial—and that we all (reasonably accurately, in our own way) understand and accept the tradeoffs along the frontier of the achievable (certainly we will, and to a certain extent should, continue to argue about the precise terms of the tradeoffs and location of the frontier though), then we can also perhaps agree that the particular vector one takes to that frontier is more a matter of taste, and that good and productive work is defined by its proximity to, and perhaps furthering of, that frontier rather than the vector it chooses. *A* probably needn't worry so much what vector *B* chooses; if *B* pushes the frontier out along her particular vector, then *A* can get further along his vector and vice versa. That is, this rosy scenario would obtain *if* we have some means of communicating, a common goal, and perhaps some common understanding of/standards for progress, toward which end hopefully this chapter has been of some utility to at least one researcher other than its author.

References

Achen, C. 1985. Proxy variables and incorrect signs on regression coefficients. *Political Methodology*, 11 (3–4): 299–316.

—— 2002. Towards a new political methodology: microfoundations and ART. *Annual Review of Political Science*, 5: 423–50.

Almond, G. 1956. Comparative political systems. *Journal of Politics*, 18 (3): 391–409.

—— and Verba, S. 1963. *The Civic Culture*. Princeton: Princeton University Press.

Beck, N. and Katz, J. N. 1995. What to do (and not to do) with time-series cross-section data. *American Political Science Review*, 89: 634–47.

Bennett, A. and Elman, C. 2006. Complex causal relations and case study methods: the example of path dependence. *Political Analysis*, 14: 250–67.

Berger, S., ed. 1981. *Organizing Interests in Western Europe*. Cambridge: Cambridge University Press.

Boix, C. 1998. *Political Parties, Growth, and Equality*. Cambridge: Cambridge University Press.

Brady, H. and Collier, D. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, Md.: Rowman & Littlefield.

Brambor, T., Clark, W. R. and Golder, M. 2006. Understanding interaction models: improving empirical analyses. *Political Analysis*, 14 (1): 63–82.

Cameron, D. 1984. Social democracy, corporatism, labor quiescence and representation of economic interest in advanced capitalist society. Pp. 143–78 in *Order and Conflict in Contemporary Capitalism*, ed. J. Goldthorpe. Oxford: Clarendon Press.

Cox, G. 1997. *Making Votes Count*. Cambridge: Cambridge University Press.

Dahl, R. 1971. *Polyarchy*. New Haven: Yale University Press.

Deutsch, K. 1971. Social mobilization and political development. Pp. 384–401 in *Political Development and Social Change*, ed. J. Finkle and R. Gable. New York: Wiley.

Dodd, L. 1976. *Coalitions in Parliamentary Government*. Princeton: Princeton University Press.

Duncan, D. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.

Franzese, R. J., Jr. 1999. Partially independent central banks, politically responsive governments, and inflation. *American Journal of Political Science*, 43 (3): 681–706.

—— 2002a. Electoral and partisan cycles in economic policies and outcomes. *Annual Review of Political Science*, 5: 369–421.

—— 2002b. *Macroeconomic Policies of Developed Democracies*. Cambridge: Cambridge University Press.

—— 2002c. Multiple hands on the wheel: empirically modeling partial delegation and shared control of monetary policy in the open and institutionalized economy. Department of Political Science, University of Michigan—Ann Arbor.

—— 2005. Empirical strategies for various manifestations of multilevel data. *Political Analysis*, 13 (4): 430–46.

—— 2006. Models for time-series-cross-section data. Lectures at Academia Sinica, Taipei, Taiwan.

—— and Hays, J. C. 2006. Spatio-temporal models for political-science panel and time-series-cross-section data. Paper presented at the 2006 Summer Meetings of the Political Methodology Society, Davis, California. **http://polmeth.wustl.edu/retrieve.php?id=626**.

—— and Nooruddin, I. 2002. Geographic and partisan bases of representation: distributive politics and the effective number of constituencies. Department of Political Science, University of Michigan—Ann Arbor.

Garrett, G. 1998. *Partisan Politics in the Global Economy*. Cambridge: Cambridge University Press.

Hall, P. 1986. *Governing the Economy*. Oxford: Oxford University Press.

—— 2003. Aligning ontology and methodology in comparative research. Pp. 333–72 in *Comparative Historical Analysis in the Social Sciences*, ed. J. Mahoney and D. Rueschemeyer. Cambridge: Cambridge University Press.

—— and Soskice, D., eds. 2001. *Varieties of Capitalism*. Oxford: Oxford University Press.

Huntington, S. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Lincoln: University of Oklahoma Press.

Inglehart, R. 1990. *Culture Shift in Advanced Industrial Society*. Princeton: Princeton University Press.

Ip, E. 2001. Visualizing multiple regression. *Journal of Statistics Education*, 9 (1): **www.amstat.org/publications/jse/v9n1/ip.html.**

Jackson, J. E. 1996. Political methodology: an overview. Pp. 717–48 in *A New Handbook of Political Science*, ed. R. Goodin and H.-D. Klingemann. Oxford: Oxford University Press.

Kam, C. D. and Franzese, R. J., Jr. Forthcoming. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor: University of Michigan Press. **www.press.umich.edu/titleDetailDesc.do?id=206871**.

Kedar, O. 2002. Policy balancing in comparative context: institutional mediation of voter behavior. Ph.D. dissertation. Cambridge, Mass.: Harvard University.

King, G., Keohane, R. and Verba, S. 1994. *Designing Social Inquiry*. Princeton: Princeton University Press.

Lehmbruch, G. and Schmitter, P., eds. 1982. *Patterns of Corporatist Intermediation*. Beverly Hills, Calif.: Sage Publications.

Lijphart, A. 1971. Comparative politics and the comparative method. *American Political Science Review*, 64 (3): 682–93.

—— 1984. *Democracies*. New Haven: Yale University Press.

—— 1994. *Electoral Systems and Party Systems*. Oxford: Oxford University Press.

Lipset, S. M. 1960. *Political Man*. Garden City, NY: Doubleday Press.

Londregan, J. B. and Poole, K. T. 1990. Poverty, the coup trap, and the seizure of executive power. *World Politics*, 42 (2): 151–83.

Ordeshook, P. and Shvetsova, O. 1994. Ethnic heterogeneity, district magnitude, and the number of parties. *American Journal of Political Science*, 38 (1): 100–23.

Political Analysis. 2005. Special Issue: Multilevel modeling for large clusters. 13 (4).

—— 2006. Special Issue: Causal complexity and qualitative methods. 14 (3).

Powell, G. B., Jr. 1982. *Contemporary Democracies*. Cambridge, Mass.: Harvard University Press.

Przeworski, A. and Teune, H. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley.

—— Alvarez, M. E., Cheibub, J. A. and Limongi, F. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990*. Cambridge: Cambridge University Press.

Putnam, R. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.

Rae, D. 1967. *The Political Consequences of Electoral Laws*. New Haven: Yale University Press.

Riker, W. 1962. *The Theory of Political Coalitions*. New Haven: Yale University Press.

Rogowski, R. 2004. How inference in the social (but not the physical) sciences neglects theoretical anomaly. Pp. 75–84 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. H. Brady and D. Collier. Lanham, Md.: Rowman & Littlefield.

Sartori, G. 1976. *Parties and Party Systems*. Cambridge: Cambridge University Press.

Smith, A. 2004. *Election Timing*. Cambridge: Cambridge University Press.

Smith, G. 1972. *Politics in Western Europe*. New York: Holmes & Meier Publishers.

Swank, D. 2002. *Global Capital, Political Institutions, and Policy Change in Developed Welfare States*. Cambridge: Cambridge University Press.

Tsebelis, G. 2002. *Veto Players: How Institutions Work*. Princeton: Princeton University Press.

Wallerstein, M. 2000. Trying to navigate between Scylla and Charybdis: misspecified and unidentified models in comparative politics. *APSA-CP: Newsletter for the Organized Section in Comparative Politics of the American Political Science Association*, 11 (2): 1–21.