

**Michael L. Ross**  
**UCLA Department of Political Science**  
**October 15, 2002**

**Testing Inductively-Generated Hypotheses With Independent Data Sets**  
*APSA-CP Newsletter, forthcoming*

Scholars in comparative politics often fail to conduct true out-of-sample tests. Many instead use a single data set to both generate hypotheses and “test” them. Of course, the resulting “tests” are not tests at all, but mere descriptions of patterns in our data. Hypotheses that are generated inductively from one data set can only be tested with a different data set. The problem afflicts both qualitative and quantitative studies.

There are many strategies for avoiding this problem. Qualitative researchers who use one case study to generate a hypothesis can conduct new case studies to test it. Alternatively, they can look for new data within their existing case – perhaps at another level of analysis – to test one of their hypotheses’ observable implications.<sup>1</sup> Quantitative researchers can seek out data sets that cover new countries or new time periods; or if they are already employing global data sets, they can subdivide them to approximate out-of-sample estimations (this latter approach – dubbed “the Stanford test” is explained in Gary King and Langche Zeng, “Improving Forecasts of State Failure,” *World Politics*, July 2001).

Of course, these strategies may be costly and are seldom employed. Too often scholars use an alternative strategy: they disguise their inductive hypotheses as deductive ones, and use the same case studies, or quantitative data sets, that they employed to produce the hypothesis to falsely “test” it. As a result, our field is strewn with hypotheses that have been dressed up as findings. As the saying goes, “it’s not what you don’t know that kills you, it’s what you know that ain’t so.” Much that we “know” in comparative politics probably ain’t so because we fail to carry out valid out-of-sample tests.

One reason these disingenuous practices are common is that we place an absurdly high value on generating original hypotheses. Many of us want to want to be theoretical innovators; few are satisfied to test the theories of others. Yet there are plenty of good, untested theories – or at least, hypotheses – waiting to be tested in the world of comparative politics, and testing them with independent data sets has many advantages.

This is what I did in my recent article, “Does Oil Hinder Democracy?” (*World Politics*, April 2001). The article takes a claim that is common among Middle East area specialists: that oil wealth inhibits the rise of democracy. It discusses the argument’s origins, reviews the case-study literature, and then tests it with pooled time-series cross-sectional data for 113 states between 1971 and 1997. In short, it finds that Mid East specialists were right: a state’s reliance on oil exports *does* tend to make it less democratic, both in the Middle East and elsewhere. It also finds that one of the theory’s implications – that other types of mineral exports will have a similar effect – is valid. Several causal mechanisms seem to account for this pattern.

I was drawn to the arguments of the Middle East specialists (even though, by training, I am a Southeast Asianist) since I am interested in the problems of resource-rich

---

<sup>1</sup> I wish to thank Dan Posner for suggesting this.

states generally, and these scholars had thought carefully about issues that interested me. At the same time, I also learned to tread carefully. The first time I presented a draft of my paper at a conference I was bombarded by criticism for idly speculating about the influence of Islam on democracy.

Still, drawing on the hypotheses of area specialists had many advantages. Area specialists commonly harbor a wealth of generalizable insights about politics, insights that are often overlooked by those of us trained in formal modeling or statistical analysis. Many of us have learned these skills because we believe they will help our work become more rigorous, and help the discipline become more scientific. If we believe in the cumulation of knowledge, however, it is absurd to ignore the scholarship others even if it has not been expressed in formal or statistical language.

Insights that are incubated by area studies specialists often have implications that reach far beyond the region's boundaries. Hypotheses about the oil-rich "rentier states" developed by Middle East specialists had important implications for oil-rich and mineral-rich countries in other regions; they even had implications for our understanding of democratization. Yet they were rarely discussed outside the tiny world of Mid East specialists.

I came to realize that the field of Middle East studies had been marginalized by political science, in ways that other area studies fields had not. Some of the most influential recent studies of democratization – including projects led by O'Donnell, Schmitter, and Whitehead, Diamond, Linz, and Lipset, and Inglehart – have avoided the Middle East entirely. Even the methodologically-astute scholarship of Pzeworski, Alvarez, Cheibub, and Limongi simply drops the oil-rich Middle East countries from the database, as though these states occupied a different universe with a unique set of physical laws. In most of these cases the methodological justification for ignoring the Middle East was weak.

A second advantage of this approach is that the results of any tests are likely to be interesting even if they are negative. When we test hypotheses that are original and generated through a deductive process, negative findings are hard to publish: after all, why should others care about our skillfully-generated hypotheses if they are wrong? Testing hypotheses that have already been produced and explored by others, however, is intrinsically interesting: either falsifying or confirming them constitutes scientific progress.

One excellent example is José Cheibub's article "Political Regimes and the Extractive Capacity of Governments" (World Politics, April 1998). Cheibub builds his article on a common argument: that a government's regime type influences its ability to raise taxes. He shows that theoretical arguments on regime types and taxes – largely produced by Latin America specialists – are often contradictory: some suggest that democracies will extract more taxes, others that authoritarian governments will extract more taxes. His own statistical analysis, based on 108 countries between 1970 and 1990, offers a true out-of-sample test of competing hypotheses. It finds that regime type has no influence on a government's extractive capacity. Even though this is a negative finding, it is still of broad interest because it rejects untested hypotheses that had been widely circulated.

A third advantage of this type of approach is that you can draw on a wealth of empirical insight when it is time to specify your model. My article drew upon a long

series of insightful case studies and theoretical innovations by Eva Bellin, Kiren Chaudhry, Jill Crystal, Terry Lynn Karl, Dirk Vandewalle, and at least a dozen others. Many had developed and illustrated hypotheses about causal mechanisms that linked oil to authoritarianism. I was able to specify and test the three mechanisms that were most widely-discussed.

Of course, understanding these case studies also entailed a lot of work. Often these scholars focused on oil's impact on a variety of dependent variables simultaneously, including regime type, economic performance, and state formation; I had to extract the oil-impedes-democracy argument from a series of wider claims about oil's political effects. Another problem was that there was no consensus among scholars on how the key variables should be defined, how different causal mechanisms operated, and to what domain of cases the oil-impedes-democracy claim should apply. The original concept – which was never terribly precise – had grown fuzzier over time, and sometimes teetered on the edge of being unfalsifiable.

One reason that the oil-impedes-democracy claim had grown stale was that there was little variation on either oil wealth or regime type inside the Middle East region. Another was that almost all Middle Eastern states face at least two other common obstacles to democracy – Islam, and a tortuous colonial history – making it hard to disentangle the partial effects of oil wealth. Together, these factors made the Mid East (ironically) an unusually *poor* setting for drawing inferences about natural endowments and regime types.

A global data set offered much more variation, and hence, a much stronger test of the claim; it also provided a true out-of-sample test. Since I wished to speak to both Mid East specialists and scholars of comparative democratization, it was important for me to correctly specify my model – that is, to find ways to statistically account for the factors that, according to case studies, had influenced the regime types of oil-rich states. By using a variable that measured the fraction of the Muslim population, I was able to control for the effects of Islam; by using dummy variables for the Persian Gulf states, and for the Middle East and North Africa, I was able to control for any region-specific effects, such as colonial history. Even with these and other controls in place, oil significantly reduced the likelihood of democracy.

Comparative politics is hip-deep in causal arguments that have never been subjected to out-of-sample tests. Many have been produced by area specialists through an “inductive” process – which often entails careful field work, historical knowledge, and innovative theorizing. Scrutinizing and testing some of these arguments is not only a good research strategy, it can also foster dialogue between qualitative and quantitative approaches, and area studies and transregional approaches, to comparative politics.