

LIMITS AND ALTERNATIVES TO MULTIPLE REGRESSION IN COMPARATIVE RESEARCH

Michael Shalev

This paper criticizes the use of multiple regression (MR) in the fields of comparative social policy and political economy and proposes alternative methods of numerical analysis. The limitations of MR in its characteristic guise as a means of hypothesis-testing are well known. The emphasis here is on the specific difficulties of applying MR to the problem of explaining diverse outcomes across a limited range of country cases. Two principal conclusions will emerge. First, even though technical means are available to deal with many of the limitations of MR, these solutions are either unconvincing or else require such advanced technical skills that they offer questionable returns on scholarly investment. Second, dissatisfaction with MR does not necessarily mandate radical alternatives or abandonment of numerical methods altogether. “Low-tech” forms of analysis (tabular and graphical methods) and multivariate statistical techniques other than MR (such as factor analysis) constitute viable and useful alternatives.

The comparative study of welfare states is a good example of the characteristic methodological polarization that afflicts the social sciences. Historians and social policy analysts with an intrinsic interest in welfare states engage in descriptive and prescriptive studies, while at the other extreme are “hard-nosed” social scientists who regard the welfare state essentially as a

Capitalisms Compared
Comparative Social Research, Volume 24, 261–308
Copyright © 2007 by Elsevier Ltd.
All rights of reproduction in any form reserved
ISSN: 0195-6310/doi:10.1016/S0195-6310(06)24006-7

convenient source of data for testing abstract theoretical claims. The sociologists and political scientists who began studying social policy in the late 1970s were part of the quantitative revolution in comparative studies. Using simple correlation and regression analysis, they optimistically hoped to settle the competition between a handful of master explanations for variation in the size of welfare states (Amenta, 1993; Shalev, 1983). Over the last two decades there has been a compelling trend toward greater sophistication in quantitative work (for a pioneering compilation see Janoski & Hicks, 1994). Especially noteworthy is the growing recognition by comparativists of the limitations of simple cross-sectional uses of MR, and their attempts to overcome these limitations without sacrificing the power of regression. Indeed, refined data analysis is the hallmark of a new and statistically more literate generation of scholars (see particularly the series *Cambridge Studies in Comparative Politics* including works by Boix (1998), Garrett (1998), Iversen (1999), Franzese (2001) and Swank (2002)). At the center of these studies are complex analyses of pooled datasets that cover multiple countries at multiple moments in time.

Earlier works in comparative political economy tended to focus on explaining enduring cross-national differences (more rarely, they looked at differences between countries in historical dynamics). The standard tools of the trade were scatter-plots, correlations and primitive cross-sectional regressions (e.g. Tufte, 1978; Cameron, 1984). This was true even of methodologically advanced practitioners (e.g. Hibbs, 1978; cf. Shalev, 1979b). The turning point was a controversial cross-national regression study by Lange and Garrett (1985) which sought to show that the combination of strong unions and left governments was beneficial for economic growth following the first “oil shock”. In a final response to their critics Garrett and Lange (1989) suggested that the debate could only be resolved by the use of a pooled cross-sectional time series design, which in addition to furnishing a much larger number of observations would enable researchers to directly study whether the effects of changes in government composition are conditioned by national institutional contexts. Two years later Alvarez, Garrett, and Lange (1991) published their seminal article “Government Partisanship, Labor Organization, and Macroeconomic Performance” which turned pooled regression into the design of choice for quantitative comparative political economists.

Alternative approaches include Ragin’s (1987, 2000) innovative attempts to formalize the analytical approach of traditional comparative-historical scholarship, and Berg-Schlosser’s demonstrations of alternative multivariate techniques (e.g. Berg-Schlosser & De Meur, 1997; Berg-Schlosser, 2002).

However, especially in the United States these methods have had little impact.¹ So far the only significant qualification to the dominance of MR in general and pooled models specifically in quantitative work on comparative political economy, has been the insistence of some practitioners on the necessity for constructive dialog between comparative history and multicountry regression analysis (see especially Hall, 2003). John Stephens and his collaborators have been the most committed exponents of this approach (Rueschemeyer, Huber-Stephens, & Stephens, 1992; Huber & Stephens, 2001), although case studies also play a subsidiary role in several notable applications of pooled regression (e.g. Boix, 1998; Iversen, 1999; Swank, 2002). Perhaps the most telling symptom of the hegemony of regression in quantitative comparative research is Gøsta Esping-Andersen's (1990) seminal work on welfare state regimes. It is striking that after offering a forceful critique of the core assumptions of conventional methodology, Esping-Andersen himself turned to MR in order to assess the empirical validity of his arguments.

The final section of this paper reanalyzes Esping-Andersen's data using techniques better suited to his theoretical and methodological premises. The preceding section offers an extended critique of pooled regression analysis. Prior to these two parts of the paper I first present an overview of the deficiencies of MR as a tool of macro-comparative research and then offer two detailed illustrations of how standard applications of MR in comparative research can generate misleading results that are inferior to those obtained using simpler methods.

STRENGTHS AND WEAKNESSES OF MULTIPLE REGRESSION

The difficulties that MR poses for comparativists were anticipated 40 years ago in Sidney Verba's essay "Some Dilemmas of Comparative Research", in which he called for a "disciplined configurative approach ... based on general rules, but on complicated combinations of them" (Verba, 1967, p. 115). Charles Ragin's (1987) book *The Comparative Method* eloquently spelled out the mismatch between MR and causal explanation in comparative research. At the most basic level, like most other methods of multivariate statistical analysis MR works by rendering the cases invisible, treating them simply as the source of a set of empirical observations on dependent and independent variables. However, even when scholars embrace the analytical purpose of generalizing about relationships between variables, as opposed

to dwelling on specific differences between entities with proper names, the cases of interest in comparative political economy are limited in number and occupy a bounded universe.² They are thus both knowable and manageable. Consequently, retaining named cases in the analysis is an efficient way of conveying information and letting readers evaluate it.³ Moreover, in practice most producers and consumers of comparative political economy are intrinsically interested in specific cases. Why not cater to this interest by keeping our cases visible?

Different views of causality are an equally celebrated source of the debate between case-oriented and variable-oriented researchers. Andrew Abbott (1998, p. 183) has cogently argued that “all too often general linear models have led to general linear reality, to a limited way of imagining the social process”. Abbot notes the constricted theoretical scope of the notion of causality underlying linear models, which cannot recognize (or at least is unlikely to recognize) situations where the effect of any given causal variable is uneven, contradictory (dialectical), or part of a wider bundle of factors sharing an elective affinity. In the social world effects are typically contingent upon their setting, including two types of historical contingency: temporal context (period effects) and time paths (particular historical sequences or cumulations). The problem is not that MR does not have or could not invent technologies for dealing with such complexities. Non-linear functional forms, interaction effects and (in time series analysis) complex lag structures immediately come to mind. The point is that because such techniques are either difficult to employ or impose a steep statistical penalty due to the “small-n problem”, they are rarely or insufficiently used.

Case-oriented analysis easily accommodates the nuances that concern Abbot and likeminded critics, because it assumes from the outset that the effect of any one cause depends on the broader constellation of forces in which it is embedded (“conjunctural causation” in Ragin’s words). If MR models try to emulate this assumption they are likely to quickly exhaust available degrees of freedom. MR is even more challenged by another causal assumption that flourishes in case-oriented analysis, namely that there may be more than one constellation of causes capable of producing the phenomenon of interest. That is, some cases are explained by one causal configuration and others by a different configuration. Statisticians refer to the phenomenon of multiple pathways to a common outcome as causal heterogeneity. MR models cannot handle this simply by increasing the number of independent variables. The results will be ambiguous because they will be unable to distinguish between additive effects, conditional relationships and multiple causal pathways.

The difficulty may be illustrated by a well-known finding of comparative welfare state research. Two subtypes of European welfare states that developed under different political auspices – Social Democracy and Christian Democracy – are known to be high spenders (for landmark studies, see Korpi, 1983; Van Kersbergen, 1995). This presents no problem for the standard additive regression model provided that the two effects are equivalent and unrelated – if for instance a strong social-democratic party could be expected to have the same effect whether or not it governed in coalition with a Christian-Democratic party. However the Austrian experience suggests that this is unlikely since historically, the black half of the “red-black” coalition severely constrained its welfare state development (Esping-Andersen & Korpi, 1984). This suggests the need for an interactive (conditional) model.

A more radical challenge to the linear additive model is posed by Esping-Andersen’s (1990), later claim that Christian-Democratic welfare states have both a policy logic and a political logic that are qualitatively different from those of Social Democracy. Although in terms of overall expenditure both social policy regimes are relatively costly, they represent two different causal syndromes that in respect to expenditure happen to result in similar outcomes. The standard regression model would treat the two political constellations as two independent variables and force them to compete to explain variance in the dependent variable. As a result the real effect of both would be diluted. And what of the hybrid Austrian case? In practice, except for the liberal English-speaking nations nearly all of the advanced political economies tend to be *either* Christian-Democratic or Social-Democratic. The peculiarities of Austrian social policy should thus be understood as the result of this cohabitation and its particular historical sequencing. They cannot be represented causally by summing the effects of the two political trends (additive model), or by trying to infer from the singular Austrian experience a law-like effect of their juxtaposition (interactive model).

To appreciate why MR is a problematic choice for comparativists, it is also helpful to consider why it may be a good choice for certain other kinds of social scientists. Economists are often interested in estimating the marginal effect of one economic variable on another, holding constant the impact of other presumed causes. If prices rise, what will be the likely effect on economic growth, net of other known influences like the rate of investment and the terms of trade? If people invest in a college degree, what will be the likely effect on their future income stream, net of other known influences like work experience? MR suits this project well. Estimating marginal effects under conditions of *ceteris paribus* is precisely what it aims to do.

In contrast, much of the curiosity of comparative political economists revolves around the presence or absence of certain conditions. Will economic growth be higher in the presence of corporatist trade unions (or a hegemonic social-democratic party, or an independent central bank)? It would be nice to know *how much* growth results from *how much* corporatism, but our theoretical interests are typically far more elementary and our predictions quite imprecise.

The evaluation of marginal effects in macro-comparative research is also dogged by the ambiguity of many of the variables of interest and the difficulty of measuring them precisely.⁴ Concepts like corporatism are so contentious that even categorical measures exhibit worrying inconsistencies (Kenworthy, 2001; Shalev, 1990). Some theoretical approaches in comparative politics are almost immune to successful quantification. An example is state-centered theory (e.g., Weir & Skocpol, 1985). Although the problem may partly be theoretical slipperiness, only superficial aspects of the structure of states (such as constitutional provisions) have proven to be measurable (e.g. Huber, Ragin, & Stephens, 1993). The framing of political action and agendas by state capacities, policy legacies and the autonomous initiatives of state managers has not been given serious consideration except in non-formal historical research.⁵ In contrast, naturally continuous variables like “left party cabinet representation” can be measured precisely. Unfortunately, however the use of such measures is rife with problems of both reliability and validity. Inter-country comparisons of long-term differences in left party power are plagued by the difficulty that, for example, a mean fraction of 50% of cabinet seats is consistent with either intermittent left government, stable left participation in cabinet coalitions, or a dominant left party which is unseated in midstream. Comparison over time is equally problematic, since the numbers alone cannot tell us whether the left’s role in government has shifted between *qualitatively different conditions* like one-party dominance, wall-to-wall coalitions, junior partnership, pivot party facing a divided right, etc. MR could accommodate such complexity by replacing the continuous measure of left strength with a series of dummy variables, or perhaps by finding an appropriate non-linear functional form to capture discontinuities in the effect of left strength on the phenomenon of interest. But the first solution is “wasteful” of precious degrees of freedom and the second requires either good luck or an unlikely degree of theoretical sophistication.

In the behaviorist sub-fields of political science and related disciplines much of the appeal of MR derives from its comfortable fit with sample survey methodology. Because they enjoy a relatively high ratio of cases to

variables, survey researchers are able to use MR as a means of introducing statistical controls. Unlike economists they may not be motivated by an ontological view that is inherently marginalist. They use controls in the hope of dealing with causal forces that in the ideal experimental design would have been neutralized by random assignment of subjects to differential “treatments”. This approach has been the subject of vigorous debate. In different ways David Freedman (1991) and Stanley Lieberman (1985) have made compelling arguments that proper statistical control would require much more sophisticated and complete causal theories than social researchers can hope to have.⁶ Even assuming that comparative political economists had such theories, given the small number of cases included in their empirical research it is technically difficult for them to analyze the effect of more than a few independent variables at a time.

Staying with the survey researchers, we can identify a final reason why the appeal of MR outside of comparative research need not inspire its use within the field. To economize on resources, analysts of voter opinion or social mobility usually poll only a tiny fraction of their target population. As a result, a fair amount of the immense heterogeneity that characterizes a universe like “American voters” cannot possibly be captured in the typical sample of only one or two thousand. Nevertheless, even the most unlikely combinations of the independent variables probably do exist in the target population. From this viewpoint one of the advantages of MR is that using the observations in hand, its coefficients (marginal effects) project relationships across the whole spectrum of potential configurations of variables.

In cross-national quantitative research the situation is very different. We often analyze the entire universe of cases, and if not it is usually because of lack of data rather than sampling considerations. For the most part then, *if a particular configuration of attributes does not exist in a cross-national dataset, it does not exist at all*. To grasp the size of the problem, consider the following hypothetical example using only three independent variables and a crude level of measurement. Social security expenditure as a proportion of GDP is regressed on left party power, exposure to trade and proportion of the population over 65. All variables are measured on a 5-point scale. If we were to construct a multiway table with this dataset, it would have 625 ($5 \times 5 \times 5 \times 5$) cells. Since no study of the OECD area can have more than about 20 cases, this implies over 600 empty cells! MR in effect places imaginary countries in some of these empty cells when it seeks out the best linear fit that can be generated for the data at hand.⁷ Because it estimates partial parameter effects as if all (linearly-fitting) configurations were possible, MR can easily yield problematic results.

The venerable social-democratic model of the welfare state illustrates this problem (Shalev, 1983). Andrew Martin's (1973) pioneering comparison of the US and Sweden inferred that social-democratic party dominance was the crucial difference responsible for Sweden's postwar commitment to the full-employment welfare state, compared with its glaring absence in the US. Numerous correlation and regression studies echoed this argument and went on to seemingly confirm its veracity across the whole spectrum of advanced capitalist democracies. Yet, this model could tell us little or nothing about the causes of policy variation between the US and other liberal political economies, or within the US over time. The coefficient for social-democratic rule generated by cross-sectional regressions yielded absurd inferences along the lines that with one additional decade of socialist rule, America (or a country like it) would probably boast an unemployment rate three points lower and child allowances 40% higher. This is an extreme example of the dangers of generalizing from empty cells when each of our cases is a complex historically bounded *gestalt*. Still, it cannot be denied that one of the tests of a useful causal model is that it will be capable of answering counterfactual questions – that is, of filling empty cells with hypothetical data. Indeed, it was precisely by asking how US policy would have developed under Swedish conditions that Martin and others were led to focus on the causal role of labor movement strength. However, some “cells” are so unlikely ever to be filled that they should not be part of either our computational space or our predictions (King & Zeng, 2002). The attributes of societies are not subject to infinite variation in unlimited combination with one another.

From an MR perspective, the problem of empty cells may not be intractable. If a variable capable of explaining differences between Sweden and the US offers no guidance to the contrast between Canada and the US, then our model must be either under-specified or mis-specified. If the problem was under-specification the appropriate response would be to add independent variables capable of accounting for the observed variation. But with these additional variables in the model, it might become too large to estimate on a small cross-sectional dataset. In response, we might be tempted to enlarge our dataset by combining cross-sectional observations for different years. This would have the added advantage of permitting the investigation of intra-country differences (i.e. within the US as well as between the US and other countries). As noted, this pooling strategy is the subject of a later section of the paper.

If mis-specification is the problem then the solution would be to find an explanation sufficiently general that it could accommodate a wider range of variation – between the US and Canada as well as vis-à-vis Sweden.

In contrast, comparativists steeped in the case-oriented tradition would be more likely to assume causal heterogeneity. Instead of looking for a new master explanation they would seek an additional one tailored to cases that are inconsistent with prevailing theory. Following this logic, in the comparative study of political economy and public policy it has become common to assume that distinctive causal trajectories apply to different “families of nations” (Castles, 1993). If MR is obviously not the best way of testing plural explanations, what is? This issue will be discussed later in the context of Esping-Andersen’s claim that there are three distinctive welfare state regimes.

Before proceeding to the questions of whether pooling resolves the problem of “too many variables and not enough cases” and whether regression is capable of dealing with causal heterogeneity, the paper offers two specific examples of the everyday use of MR. These illustrations were chosen with an eye to countering two possible responses to the general critique of MR that has been offered so far. One of these would be to lower our expectations and utilize regression more as a means of partitioning empirically observed variance than of rigorously testing hypothesized causal relationships. Alternatively, it might be argued that the causal status of regression coefficients should indeed be treated tentatively, but that our confidence is strengthened if alternative types of numerical and non-numerical analysis yield convergent findings. Both approaches have their problems. The next section critiques an illustration of the use of MR as only a loose guide to the plausibility of alternative models. Using a different example, the section that follows shows that even convergence among different methodologies does not guarantee that the data will yield their fundamental secrets.

“CAUSAL ARGUMENTS” OR MERE “SUMMARIES”?

With multidimensional data sets, regression may provide helpful summaries of the data. However, I do not think that regression can carry much of the burden in a causal argument. (Freedman, 1991, p. 292)

David Freedman is a statistician who believes in the power of numbers but has made it his mission to disabuse social scientists of their exaggerated belief in statistical inference as a tool of causal analysis (Freedman, 1985, 1987, 1991). The essence of the argument made by Freedman (see also Leamer, 1983) is that statistical hypothesis-testing requires that researchers have a well-developed theory and a hands-off relationship with the data

prior to the point at which testing is carried out. In practice social-science research is based on weak or incomplete theories and its empirical generalizations are almost always the outcome of numerous iterations. Accordingly, when forced to confront the fact that progress in social research rests on a “dialog of ideas and evidence” (Ragin, 1994b), one should concede that the most which can legitimately be done with MR is to use it to summarize multivariate datasets.

Given prevailing expectations regarding publishable research, few scholars have the courage to claim that their research objectives are purely descriptive (Abbott, 1998). Still, some comparative research has treated MR as less than a formal hypothesis-testing device and more like an economical method of sustaining broad empirical claims. An example of this low-expectations approach can be found in Rothstein’s (1990) study of cross-national variation in union membership from a new institutionalism perspective. Although Rothstein’s article was primarily based on comparative-historical analysis, it included a simple cross-country regression. The substantive background to the study was that under the so-called “Ghent system” unions bear responsibility for administering unemployment insurance, with the consequence that in periods of economic crisis or transformation their membership is unlikely to be eroded and may even increase. For theoretical reasons, Rothstein wished to demonstrate that the highest levels of unionization have been reached only in countries where this system is in place. His union density figures for 18 OECD countries in the mid-1980s reveal that Ghent is indeed present in all of the countries with the highest rates of union penetration, and only these countries. Hence, unless Ghent is but a spurious understudy for the real star of the causal show, it has been a necessary condition for rates of more than 70% unionization. Of course, this does not mean that the Ghent system is a *sufficient* condition for union success. Perhaps it merely amplifies the effects of other favorable conditions.

There are thus several possibilities that a simple table showing union membership alongside Ghent presence/absence cannot address: spurious association (alternative explanations), additional causes (complementary explanations), and interaction effects (conditional explanations). Following convention, Rothstein seeks to lay the first two of these issues to rest by executing a multiple regression that takes into account other probable influences on cross-country differences in unionization. These are left party participation in government, and potential union membership (the absolute number of employed and unemployed wage-earners).

Rothstein's model was re-estimated for this article using a modified version of his dataset.⁸ Following the original, the coefficients are standardized *betas*.

$$\begin{aligned} \text{Percent Unionized} = & 0.47(\text{Ghent}) + 0.28(\text{Left Government}) \\ & - 0.34(\text{Log of Potential Membership}) \end{aligned}$$

All coefficients are significant at conventional levels (although Left Government only marginally) and the adjusted *R*-squared is 0.73. The metric coefficient for the Ghent variable reveals that the net average difference in unionization between Ghent and non-Ghent systems is a striking 27-percentage points.

Notwithstanding these indications of success, it can be argued that Rothstein's use of MR is inappropriate and in part misleading. Rothstein is content, in his words, to show "that all three variables have an independent explanatory effect of about the same standardized size" (Rothstein, 1990, p. 41). However, a prerequisite for these "explanatory effects" to have causal meaning is that the model be theoretically plausible. Rothstein himself casts doubt on this, when he describes the argument for the significance of potential membership size as logically indefensible, and suggests that the left-government argument suffers from what econometricians call simultaneity bias. In addition, while the standardized coefficients indeed suggest that Ghent has at least as much empirical weight as rival explanations, because countries are invisible the results do not speak to Rothstein's core claim that it is Ghent, not left strength or small size, which differentiates between the most unionized countries and all the rest. True, this claim would have been negatively ruled out had the Ghent effect disappeared once the other variables were added to the equation. But the regression could not make a positive case for Rothstein's argument.

Beyond these specific limitations of MR in Rothstein's case, his model rests on a standard but questionable assumption. Rather than operating as a syndrome of elective affinities, the explanatory variables are assumed to exert causally distinct effects. Consequently, none of the effects is assumed to be conditional on the value of other variables – i.e. no interactions are anticipated.

A straightforward way to address these issues is to summarize causes and effects in a way that identifies different combinations of conditions (causes) with the countries that "carry" them. This requires some forethought because Rothstein's model refers to three different causal variables and his dependent variable, unionization, is not easily collapsed (it is distributed

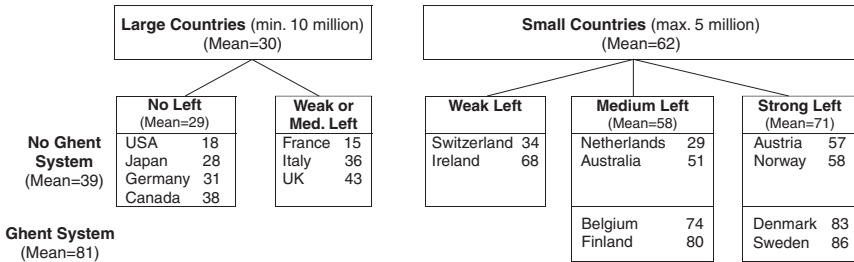


Chart 1. Reanalysis of Rothstein's Model of Union Membership.

fairly evenly across a broad spread). The proposed solution is a simple flow chart or “tree” showing exact values of unionization for different clusters of countries. These clusters were created simply by cross-tabulating the presence or absence of Ghent with categorical versions of Rothstein’s two other causal variables.⁹

The results (Chart 1) offer interesting evidence of nested causal effects. This is immediately apparent from the systematic difference between extant and non-existent configurations. Substantial left party representation was only attained in small countries, and only countries with a substantial left had the Ghent system.¹⁰ In the case of the affinity between Ghent and left strength, Rothstein himself pointed out that we cannot know which way the causal arrow points without branching into historical research. Indeed, this is true of all of the relationships among unionization, Ghent and left strength.¹¹ But we can say that c. 1985, it is the *combination* of smallness, “leftness” and Ghent that is associated with the highest rates of unionization. The results also hint at a more specific interaction. The Ghent effect may be stronger in countries with medium left strength than in the fully fledged social democracies.

This “unsophisticated” method of presenting the data reveals regularities that MR does not. In the process it more effectively vindicates Rothstein’s thesis by making clear precisely what he wanted to demonstrate: that the Ghent effect is large and not spurious, and that it comes into play in countries where other conditions are broadly favorable to unions. But these results do something else important, which is to point the interested researcher to the most fertile questions for selective case comparisons that might help nail down how important Ghent really is.¹² In particular, it must be questioned whether the Ghent system alone can explain the very large differences in density between otherwise well-matched countries: Belgium vs. the Netherlands, and Sweden and Denmark vs. Norway.¹³

The visibility of the relationship between variables and cases in the simple diagrammatic presentation favored here may thus draw attention to anomalous cases, which reveal limitations in the theoretical model. Attending to outliers from a regression analysis is sometimes also a way of identifying anomalies, but not of the kind discussed here – namely countries that do not “make sense” *when viewed in relation to other similar cases*. Tabular or graphical presentation of the dataset with named observations permits this; inspection and diagnostic testing of regression residuals does not.

COMPLEMENTING REGRESSION WITH OTHER TYPES OF ANALYSIS

Peter Hall and Robert Franzese (1998) have contributed to a significant subfield of comparative political economy which challenges the preeminence of economists in studying central banks and their impact on economic performance (Iversen, Pontusson, & Soskice, 1999). Hall and Franzese argue that while independent banks are always anti-inflationary, under certain institutional conditions their impact on the labor market is far less salutary. Unless wage setting is centralized and coordinated the bargainers will fail to internalize bank “signals”, and the result will be higher rather than lower unemployment.

In testing their argument Hall and Franzese proceed in three stages. First, they demonstrate its plausibility by referring to the paradigm case of West Germany. Second, they use data for 18 OECD countries over the entire postwar period, presented in a simplified tabular format. Finally, they use MR to test a more elaborate model at several levels of aggregation ranging from full-period means (pure cross-section) to pooled annual data. The results of each one of these analyses are consistent with their argument that the impact of central bank status on unemployment is conditional on the structure of wage bargaining.

In their initial quantitative analysis, Hall and Franzese collapse measures of central bank independence (hereafter CBI) and wage coordination and cross-tabulate them. The results clearly confirm the hypothesized interaction effect. However the authors recognize that this effect could be an artifact, the result of some confounding influence like countries’ wealth, economic openness or government composition. In practice, the result survives the application of controls for these variables using MR. Conditional parameter estimates show that the interaction between independence and coordination is substantively as well as statistically significant. Moreover, diagnostic

testing indicates that these results do not depend on the presence of any particular case.

Hall and Franzese's study deserves close attention precisely because it offers such a thorough application of MR, which moreover very sensibly builds on prior qualitative research on the German case. Yet it will be shown that the study's tabular results are misleading. Missing from these results is an element which proved crucial in probing Rothstein's study, namely, identification of the cases (countries). Another issue is how best to group continuous data into categories in order to reveal multivariate relationships. It was relatively easy to categorize Rothstein's variables intuitively, but this is not the case for Hall and Franzese's data. Although formal methods are sometimes used for this purpose (e.g. Goodman's (1981) test of "collapsability"), most researchers rely on commonsense ways of determining cutoff points: substantive familiarity with the cases, aggregation into categories of similar size or tailoring the categories to breaks in the distribution of observations. Hall and Franzese provide no explicit rationale for their cutoff points. Taking advantage of the availability of their dataset,¹⁴ Chart 2 permits direct examination of the distribution of cases along the two institutional dimensions. Visual inspection of each dimension offers no indications of categories that could be "naturally" amalgamated. Further, observing the two-dimensional patterning of the countries one is not struck by any

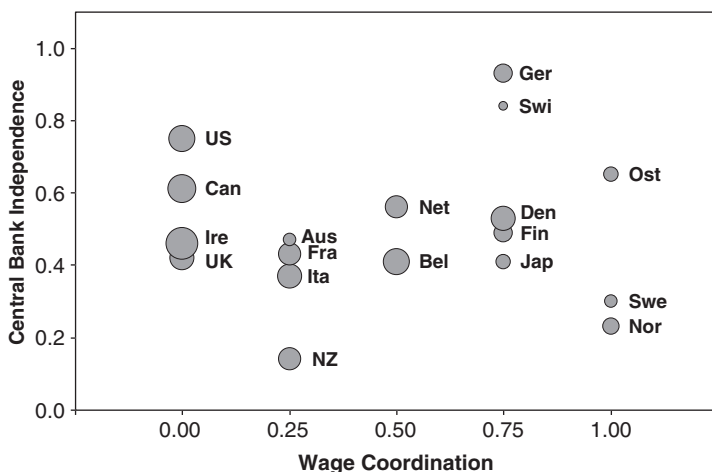


Chart 2. Institutional Configurations (X and Y axes) and Unemployment (Bubbles) (Based on Hall and Franzese).

Table 1. Institutional Effects on Unemployment
(Derived from Hall and Franzese).

Wage Coordination	Central Bank Independence	UE 1955–1990	UE 1955–1973	UE 1984–1990
0.00	Lower (UK, Ire)	6.8	4.0	12.9
	Higher (US, Can)	6.2	4.9	7.6
0.25	Lower (NZ)	4.2	2.1	7.6
	Higher (Aus, Fra, Ita)	3.9	2.3	7.5
0.75	Lower (Den, Fin, Jap)	3.3	2.0	5.3
	Higher (Ger, Swi)	2.0	0.8	4.2
1.00	Lower (Nor, Swe)	2.0	1.8	2.6
	Higher (Ost)	2.2	1.8	3.5

Source: Hall and Franzese dataset (made available at the URL cited in note 15). Differences between the average unemployment rate for 1955–1990 reported here and in Table A.1 of Hall and Franzese (1998) are due to an error in the published table (Robert Franzese, personal correspondence, November 6, 2002). *Abbreviations:* Ire, Ireland; Can, Canada; NZ, New Zealand; Aus, Australia; Fra, France; Ita, Italy; Den, Denmark; Fin, Finland; Jap, Japan; Ger, Germany; Swi, Switzerland; Nor, Norway; Swe, Sweden; Ost, Austria.

obvious clustering. This suggests that Hall and Franzese may have erred in collapsing their institutional variables into dichotomies.

Is it possible without aggregation to discern the effects on unemployment, which were apparent in Hall and Franzese’s aggregated figures (their Table 1)? The “bubbles” in our chart are proportionate in size to the mean unemployment rate for 1955–1990 in each country. Looking first for univariate effects, it is noticeable that as we move from left to right along the x-axis the jobless rate drops quite dramatically. No such clarity is evident when comparing unemployment rates at lower and higher levels of CBI (i.e. moving from the bottom to the top of the y-axis). Consequently, whereas unemployment is strongly correlated with wage centralization ($r = -0.74$) it is completely uncorrelated with CBI ($r = -0.07$).

The critical question though is whether “In nations where wage coordination is high, an increase in the independence of the central bank is associated with a very small increase in the rate of unemployment Where wage coordination is low, however, an increase in the independence of the central bank is associated with a substantial increase in the rate of unemployment” (Hall & Franzese, 1998, p. 518). Chart 2 provides no evidence for this proposition. In fact unemployment fails to rise with the extent of CBI at all levels of wage coordination. Apparently, the aggregation of Hall and Franzese’s original data into categories inadvertently generated unfounded support for their hypothesis.

There is also an important substantive issue, which their analysis fails to reckon with. Studies that pool data from different points in time – whether by simple averages or complex panel analysis – implicitly assume stability in the causal relationships under consideration.¹⁵ However, in the aftermath of the second oil shock, unemployment in most European economies rose dramatically while in North America it declined. Was this shift in international unemployment differentials, which persisted into the 1990s and beyond, accompanied by a change in the conditional impact of CBI? To find out, **Table 1** compares unemployment in the postwar golden age (defined here as 1955–1973) with the period of global crisis from 1984–1990 (when the time series ends). Given that “our key institutional variables do not vary over time” (Hall & Franzese, 1998, p. 520), no attempt has been made to calculate sub-period measures of centralization and CBI. Further, to simplify the presentation **Table 1** builds on the fact that *within each level* of wage coordination two groups of countries are discernable, one with higher CBI scores than the other.¹⁶ The table permits us to evaluate whether relatively higher levels of CBI are associated with higher unemployment as coordination declines, in both the complete series and the two sub-periods.

The results confirm that the data for the postwar period as a whole do not fit expectations, but they show that in the period prior to 1974 there is some support for the predicted conditional relationship. This support would be stronger but for the fact that the two uncoordinated economies with low CBI, Ireland and the UK, experienced very different unemployment rates. The CBI “penalty” in this period thus turns heavily on the question of whether the role of the central bank can carry the main explanatory weight for the contrast between the UK, with well under 3% average unemployment; and the US and Canada with nearly 5%. I believe that a stronger explanation is provided by the absence of social democracy in North America compared to the paramount influence of the Labour Party on the terms of Britain’s postwar settlement (Korpi, 1991). Turning to the later period of economic crisis, **Table 1** shows that the results are at odds with Hall and Franzese’s expectations. Among the least coordinated economies, North American unemployment was actually lower than in Britain or Ireland.

Perhaps one should not place too much weight on evidence concerning the gross effects of institutional context on economic performance. The authors of the study saw tabular analysis as only one building block in a longer evidentiary chain that included cross-country regressions controlling for key economic and political influences on unemployment (including the variable just referred to, government partisanship). Moreover with unusual thoroughness they ran these regressions not only on cross-sectional averages

for the entire postwar period, but also used pooled time series data in the form of either decade-long averages or annual observations. They report that the results of all of these tests were consistent with their leading hypothesis.

Nevertheless, there are reasons to take a cautious view of Hall and Franzese's multivariate analysis. With four control variables entered in aggregate cross-country regressions alongside the two institutional indicators and their interaction, the model is seriously overweight for application to only 18 cases. In theory, this limitation ought to be overcome once multiple observations for each country are combined at different time points. But for reasons that will be explicated in more detail in the next section of the paper, this is questionable. For instance, as we have just seen the postwar period 1955–1990 was far from homogeneous in its unemployment record. The models used by Hall and Franzese do control for over-time variability in the overall level of joblessness, but not for the equally plausible possibility that the determinants of unemployment altered over time.¹⁷ In addition, whether tested in sparse cross-sectional format, decade-long panels or by pooling annual time series across countries, these regression models build on a great many empty cells. The vast majority of potential combinations of collective bargaining systems, CBI, union and left party strength and trading conditions have no empirical counterparts. As in most studies of this type, multiple time frames primarily add more cases to already-populated configurations.

The implications of limited diversity in the dataset utilized by Hall and Franzese are especially worrying for their most impressive evidence – decadal averages that simulate “what difference it makes”. The authors' Table 4 presents expected levels of unemployment for 15 different institutional configurations, calculated by fixing control variables at their sample means. The results indicate that, as predicted, the effect of CBI is profoundly influenced by the degree of wage coordination. In completely uncoordinated systems unemployment is expected to be nearly *10 points higher* at maximum bank independence than at the minimum level of CBI. In completely coordinated systems there is a modest effect in the opposite direction. These results contrast very strongly with the uncontrolled effects that we have observed. However, it turns out that of the 15 cells in Hall and Franzese's table approximately two-thirds have no empirical counterparts. As it happens, the contrasts among the “extant” cells, while in the expected direction, are far more mild than those based on the hypothetical extremes of the institutional matrix.¹⁸ Moreover the predicted levels of unemployment are seriously off the mark, higher than the real ones for decentralized systems and lower for the centralized ones.

There is a possible explanation for Hall and Franzese's inaccurate predictions of unemployment levels that also casts doubt on the veracity of their simulated effects of CBI (even for the realistic configurations). Both results may be traceable to the effect of elective affinities. As noted, Hall and Franzese adopted the typical procedure for such "what-if" exercises, allowing the explanatory variables of theoretical interest to vary while controlling for additional known influences by calculating their impact at mean levels. However as already noted in connection with Rothstein's study, different elements of the institutional context tend to cohere. For instance, coordination generally thrives in small, highly unionized economies with strong social-democratic parties but is stymied in liberal political economies with the opposite set of features. Consequently, by evaluating their control variables at the grand mean for all countries it is likely that Hall and Franzese inflated their predictions for the coordinated economies and understated them for the decentralized ones. The same bias may have exaggerated the deleterious effect of CBI in the decentralized context.

To sum up, Hall and Franzese present us with a study that is impressively well-rounded methodologically, integrating qualitative and quantitative research and moving stepwise from simple to sophisticated forms of numerical analysis. Despite this, their quantitative results are unconvincing. By failing to address temporality, limited diversity and elective affinities, their multivariate analyses almost certainly overstated the potency of the effects they sought to uncover. Their tabular analysis, based on questionable category groupings and abstracted from the cases under study, generated misleading results. In small-n comparative research even an analytical device as simple as a cross-tabulation needs to be applied with close attention to the data at hand. The pitfalls of the pooled regression models used by Hall and Franzese make it clear that more complex techniques offer no guarantee of yielding an empirically plausible account. While by now these pitfalls are well known they have not deterred comparative quantitative researchers from wholesale adoption of pooled MR as their technique of choice. The next section of the paper provides a fuller account of the problems this entails.

IS POOLING A PANACEA?

Some readers might view elements of the critique of the two articles discussed so far as just another illustration of a well-known problem: that because comparativists have "too many variables chasing too few cases", MR can only be applied either crudely (Rothstein) or else implausibly (Hall

and Franzese) in standard cross-sectional designs. My alternative approach might be criticized as a dishonorable retreat to rendering descriptive summaries of the data that are all too dependent on arbitrary decisions about how to group and present them. These critics would doubtless reject my argument that regression is fundamentally unsuited to macro-comparative analysis, and would prefer to focus their creative energies directly on solving the problem of insufficient cases (e.g. King, Keohane, & Verba, 1994, pp. 24, 30–31).

In this spirit, John Goldthorpe has argued that “*au fond* the small-N problem is not one of method at all but rather of data”. Goldthorpe specifically recommends emulating the large number of researchers who “have ‘pooled’ data for the same set of nations for several different time-points. Observations – and degrees of freedom – are in this way increased ...” (Goldthorpe, 1997, p. 8).¹⁹ However, there are well-established reasons to believe that the most likely consequence of a turn to pooling is to muddy the causal waters still further. My critique proceeds in three stages. First, I explain why the rationale for using pooling as a means of adding statistical degrees of freedom is fundamentally flawed. Second, I demonstrate that creative attempts to overcome the difficulties of making causal inferences from pooled data are encouraging in principle but have been of limited practical benefit. Third, pooling encounters severe technical stumbling-blocks, and it is questionable whether growing methodological sophistication will reliably overcome these difficulties.

What does pooling entail?²⁰ Traditionally, quantitative macro-level research analyzed either “snapshots” of different countries at a single moment in time (cross-sectional data), or else period-to-period data for a single country (annual time series or sub-period averages). Pooled datasets merge these two views by “stacking” panels for multiple countries one on top of the other. Hence they embody both comparative variation between countries and dynamic variation over time. As a result analysts must contend with the technical complications characteristic of both cross-sectional and time series estimation, and practitioners face a bewildering range of technical problems and solutions. Even more basic is the well-grounded fear that pooling may be counter-productive “if thoughtful consideration is not given beforehand to the *meaning* of the aggregations in the pool” (Sayrs, 1989, p. 70).

Most comparative researchers who use pooled designs have been motivated by the traditional agenda of cross-sectional comparison, the desire to explain enduring differences between countries. These researchers implicitly regard each cross-sectional snapshot as just one more view of the same between-country variability. However, it has long been understood that the

effect of a given independent variable may be quite different in time series and cross-section “because the underlying causal structures differ” (Firebaugh, 1980, p. 333). For instance in their comparative and historical study of class conflict Korpi and Shalev (1980) observed that while temporal fluctuations in strikes followed an economic logic, with falling unemployment stimulating greater labor militancy, the cross-sectional variance followed a political logic, with lower unemployment operating as a disincentive to strong labor movements to employ the strike weapon. In this spirit, Hicks (1994, p. 171) promoted pooling precisely as a means of carrying out “systematic comparisons of cross-sectionally and longitudinally varying causal forces”. But the reality is that most pooled designs utilize multiple cross-sections in order to fortify comparative generalizations, or multiple time series to fortify dynamic generalizations, on the implicit assumption that there is no difference in causality between the two dimensions.

A quite different, and more constructive approach to pooling, is to exploit the combination of comparative and over-time data in order to uncover and explain cross-national differences in over-time processes. Examples of this type of enquiry can also be found in studies of the political economy of class conflict (e.g. Hibbs, 1976; Shalev, 1979a). Time series regressions on strike activity in different countries yielded divergent results. Some scholars saw this simply as an antidote to exaggerated generalizations (Paldam & Pedersen, 1982). But others interpreted diverse parameter estimates as exemplifying the predictable effect of contextual forces on conflict dynamics (Snyder, 1975).

This has been the tack followed by the most thoughtful analysts of pooled datasets, Larry Griffin, Larry Isaac and their associates (Griffin, Barnhouse Walters, O’Connell, & Moor, 1986; Griffin, O’Connell, & McCammon, 1989). In what is still the best exposition of pooling for comparative political economists, Griffin et al. (1986) used annual data for 12 nations and 16 years to explore the effects of six economic and political variables on countries’ expenditure on income maintenance. Their first finding was that the bulk of the variation in most of their independent variables was concentrated in *either* the time or cross-country dimension. This alone suggests that it would not have made sense to use a single model to explain both dimensions. And indeed, Griffin et al. found that “the average cross-national slopes and the average time series slopes ... have very little in common” (p. 116). Even within the time and space dimensions, the contingency of causal relations could not be ignored. The results of annual cross-sections proved to be “extraordinarily unstable across years”, even *contiguous* years (p. 111). While country-specific time series estimates were more stable, they

nevertheless seemed to “evoke markedly different processes” (p. 115). Despite these reasons not to treat pooled data simply as more data, it is rare for analysts to differentiate between over-time and cross-sectional effects or to take seriously the possibility of temporal or national specificity.²¹ True, it is not uncommon for pooled models to include dichotomous variables intended to capture country or period effects. However, what these dummies actually measure are differences in the intercept or “baseline value” of the dependent variable in different countries or years. Interaction terms, far more costly in degrees of freedom, would be required to test country or period differences in *slopes*.²²

For those mainly interested in explaining dynamic processes, on the other hand, pooling makes it possible to contemplate multiple explanations tailored to different contexts. The dynamics characteristic of a country or group of countries might be seen as both indicative of, and caused by, long-run (structural) differences. Griffin and his colleagues proposed a systematic methodology for this type of research. They suggested that time series parameters be estimated in regressions for individual countries. In a second round, these parameters would be treated as dependent variables to be explained cross-sectionally by broad-brush differences between countries (Griffin et al., 1986). While this technique may produce suggestive results (cf. Griffin et al., 1989), the credibility of the second-round results is, of course, dependent on the quality of the first round of time series estimates. Since these are typically based on short series, which may themselves be punctuated by causal heterogeneity, it is hard to be confident about these estimates.

Bruce Western (1996, 1998) has, however, offered an attractive approach to conceptualizing and estimating the type of multilevel design proposed by Griffin and his associates. Western (1996) sought to show that institutional factors like the presence or absence of corporatism could explain differences between countries in the dynamic effects of variables like government composition on fluctuations in unemployment.²³ He advocated a Bayesian approach to estimation that allows for possible contextual differences in causal dynamics, but differs in an important respect from Griffin’s two-stage method. Western’s technique permits estimates for individual countries to “borrow strength” from the whole sample. The implications of this are profound. It seemingly allows the analyst to take advantage of the more numerous observations and greater diversity afforded by pooled datasets, without having to assume identical causality in both time and space. Pooling would then be freed of most of the objections I have raised and, as Western explains, the issue of whether comparativists ought to generalize within or

beyond specific contexts would become a tractable empirical question rather than an epistemological conundrum.

Western's success in this regard is best assessed by considering the results of his own illustration, an analysis of unemployment using a pooled dataset for 18 OECD countries between 1964 and 1990 (Western, 1996). Impressively, he was able to demonstrate corporatism's implications in both the long and short run. Over the long run (cross-sectionally), corporatist countries were found to experience significantly lower rates of unemployment. From the dynamic (time series) perspective, the evidence supported the common claim that corporatism safeguards employment by improving the short-run tradeoff between wages and jobs. However, Western obtained puzzling findings for the dynamic effects of shifts in government composition. They appeared to show that in corporatist countries and other settings where collective bargaining is widespread, *increases in left party power cause unemployment to rise*. As always, the credibility of statistical conclusions needs to be checked against the cases. Chart 3 reproduces Western's estimates of the dynamic effects of changes in left cabinet representation. To highlight possible institutional consequences of the type Western was interested in, countries have been grouped using his indicators into three different settings – "unregulated", "regulated" and "corporatist".²⁴

At first sight, Chart 3 strongly confirms the finding that "social democratic governments tend to raise unemployment where collective bargaining

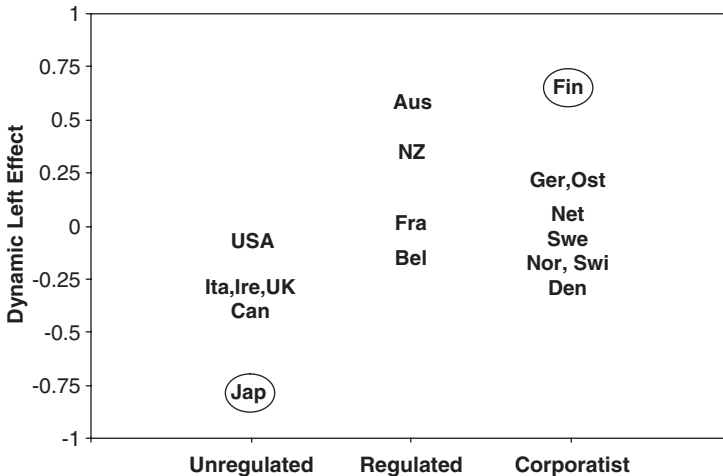


Chart 3. Western's Hierarchical Model of Unemployment.

coverage is extensive” (Western, 1996, p. 25). However without two outliers – Japan and Finland – this tendency would be substantially weaker.²⁵ As it happens, the dynamic effects of leftwing governance in these two critical cases are highly problematic. During the period studied by Western, Finland experienced few significant shifts in the left’s overall role in government. (What *did* vary was the relative role of the communist and socialist parties, a feature of government composition not measured in his study.) As for Japan, in the relevant period its left party representation was an unvarying zero.²⁶

Western’s hierarchical approach to utilizing pooled datasets holds out the possibility of harnessing their wealth of information while simultaneously respecting and even exploiting the difference between synchronic and diachronic causation. However, the key to reconciling these two objectives is “borrowing strength”. In Western’s words, “Information from other countries will help provide an estimate for a coefficient in a particular country where, say, a given independent variable shows no variation” (Western, 1998, p. 1240). This approach rests on a strong belief in the possibility of generalizing from “populated cells” to “empty cells”. In the example at hand, the dynamic effects imputed to two cases generated extreme values that became the foundation on which a strong cross-national generalization was built. It is difficult to have confidence in such a generalization. This is a pity because Western’s analytical strategy is very inviting to comparativists. Instead of merging repeated cross-sections simply in order to beef up the number of cases, he drew on the nested logic of multilevel modeling (Steenbergen & Jones, 2002). Moreover, he asked a question quintessential to the comparative method: do over-time relationships differ across countries and if so what stable differences between countries can predict those differences? Viewed this way, the pooled design offers an empirical way out of the controversy over whether causation is contextual (proper names are indispensable) or general (proper names surrender to variable names). In practice, however, since efficient estimation risks basing our ultimate conclusions on implausible counterfactual evidence, there may be no alternative to statistically unreliable country-by-country analyses.

Beyond issues concerning the analytical and practical justifications for the pooled design, as Stimson (1985, p. 945) pointed out at an early stage of the pooling revolution in political science, the technique suffers from “a plethora of potential problems” of a more technical kind. The validity of any regression estimate rests on assumptions about the statistical properties of the data, in particular the distribution of prediction errors. The characteristic problem for analysis of data collected at different time-points is serial correlation, which means that there is some kind of trend in the errors (e.g.

they tend to get bigger or smaller over time). For cross-sectional regressions comparing different units at a single moment in time, the typical challenge is “heteroskedasticity”, meaning that the errors vary with the level of a predictor variable (e.g. corporatism may be a better predictor of unemployment in more corporatist than less corporatist countries). Further, cross-sectional errors may be “locally” interdependent. Examples commonly noted in comparative political economy are policy diffusion from one country to another through bilateral or multilateral coordination, or the economic impact of big countries on their smaller trading partners. From a technical point of view, pooled designs are the worst of both worlds. They expose regression estimates to the risks of trends in the error structure over time *and* systematic variation in the error term across units. To make matters worse these problems may appear in subtle combination, for instance heteroskedasticity could increase over time. In addition, if as we have suggested explanations may have differing applicability at different moments (or periods) and across different countries (or families of countries), then the errors will also be patterned by causal heterogeneity.

There are numerous ways to shield the accuracy and reliability of regression coefficients from these risks. However, many of them are atheoretical technical fixes that treat the deviant phenomena as “nuisance” rather than “substance” (Beck & Katz, 1996). In addition, the inferences generated by different remedies are often wildly dissimilar, while at the same time it is not entirely clear which remedy is the “right” one (Stimson, 1985). So far as causal heterogeneity is concerned, our earlier discussion has shown that conventional solutions to the problem are either wasteful of degrees of freedom or require heroic assumptions concerning the transferability of relationships from one context to another.

These issues are exhaustively treated in the pedagogical literature already referenced here (Beck & Katz, Griffin, Hicks, Stimson and others) as well as in standard econometrics texts. What bears emphasis is the questionable relationship between the costs and benefits of pooling, given that its technical complexities render it a risky and uncertain enterprise and at the same time one which imposes a steep and continuously rising learning curve. Most practitioners have responded to this dilemma by looking to “best practice” and following it faithfully – often with disastrous consequences. The breakthrough article by Alvarez, Garrett, and Lange (1991) referred to earlier utilized a Generalized Least Squares technique then regarded as state-of-the-art. However Beck et al. (1993) famously showed that because their dataset included more countries than time-points, this technique gravely inflated the significance of most parameter estimates. Subsequently,

Beck and Katz (1995) demonstrated that this problem invalidates the results of numerous well-known applications of the pooled design in comparative political economy and they introduced a new technique for estimating standard errors. Beck and Katz (1996) made the further suggestion that the dynamics generating serial correlation of time series errors should be modeled by including the lagged dependent variable as a predictor.

While Beck and Katz's proposals have subsequently become virtually canonical in modeling pooled data in political science, they have been sharply criticized by some other specialists. Achen believes that under typical conditions of high serial correlation and trended exogenous variables, "the lagged [dependent] variable will falsely dominate the regression and suppress the legitimate effects of the other [independent] variables" (Achen, 2000, p. 24). Specialists in international relations (where research designs are often much less constricted in degrees of freedom) have also engaged in heated debate concerning the use of pooled models.²⁷ An eminent econometrician has characterized Beck and Katz's prescriptions as "not, strictly speaking, correct", adding that "the procedure of using OLS and reporting the 'panel corrected' standard errors is sweeping the problems under the rug" (Maddala, 1998, pp. 60–61).

One of the few critical voices heard within comparative political economy is that of a European scholar, Bernhard Kittel. After reviewing many of its technical and practical deficiencies, Kittel (1999, p. 245) concluded that pooling adds statistical value to static cross-sectional regressions only "under quite demanding conditions and to a very limited degree". A more recent contribution by Kittel and Winner (2005) offers an exhaustive replication of a typical contemporary study, by Garrett and Mitchell (2001). On the basis of numerous alternative methods of testing and evaluation it is concluded that the results of this study are empirically unfounded. An even more sophisticated dissection of the same study by Plumper, Troeger, and Manow (2005) not only reveals additional technical deficiencies, but also challenges some of the main substantive conclusions drawn by Kittel and Winner.

The level of methodological expertise required to follow these kinds of debates over pooling has become prohibitive for many scholars. In rare but encouraging instances, analysts who are not professional methodologists have questioned technical orthodoxy because it generated results that simply did not make sense. Thus, Huber and Stephens (2001, Ch. 3) rejected the use of the lagged dependent variable as a predictor of social expenditure, arguing that it would have redefined their research question from assessing the long-run impact of differing political configurations to predicting short-run

fluctuations. Indeed, given the complexity of political dynamics and the poor likelihood of capturing them by crude measures like short-run changes in the proportion of the executive controlled by social or Christian-democratic parties, it is not surprising that in study after study political partisanship loses its explanatory efficacy once the design shifts from explaining levels to explaining dynamics. (See also [Plumper et al., 2005](#); but compare [Podesta, 2003](#).)

Because available techniques are constantly updated by statisticians and econometricians, quantitative political economists are tempted to devote much time and effort to refining their skills with pooled models. There are optimists who believe that such refinements can resolve the fundamental issues raised here, but in my judgment it is more likely that our theoretical understanding of causality will continue to far outstrip our measurement and estimation capabilities. Nevertheless, it should be noted that there has recently been a mushrooming of innovative statistical methods designed to address some of the problems discussed here.

[Beck and Katz \(2003\)](#) have suggested a variety of ways to systematically assess whether pooling multilevel data is justified, and [Zorn \(2001\)](#) has proposed a method of distinguishing between dynamic and cross-sectional effects. [Braumoeller](#) has developed new techniques for incorporating central goals of [Ragin's](#) approach into the regression framework – testing for the presence of necessary and sufficient conditions and modeling causal heterogeneity ([Braumoeller & Goertz, 2000](#); [Braumoeller, 2003](#)). In a similar spirit, [Giroi and King \(2001\)](#) have devised a method of allowing explanations of over-time variation to vary across countries. But there is also bad news to report. [Braumoeller's](#) method of identifying multiple causal paths is only viable if the cases “represent all combinations of conditions” ([Bear Braumoeller, personal correspondence July 23, 2005](#)), while [Giroi and King's](#) technique seems to require a very large number of cases.

Finally, [King, Tomz, and Wittenberg \(2000\)](#) have proposed a simulation technique for increasing the amount of information on which statistical inferences are based, thereby enhancing their accuracy and certainty. [King](#) and his collaborators used this method to enthusiastically confirm a key finding of [Geoffrey Garrett's](#) influential book *Partisan Politics in the Global Economy*. Because this example poignantly illustrates the extent to which technique may outstrip data fundamentals, it deserves a closer look.²⁸

[Garrett's \(1998\)](#) aim in using pooled regressions was to assess how the distribution of class power affects policy responses to globalization. These regression results were the basis for estimating expected levels of economic performance and public spending under different political configurations,

controlling for other relevant influences. Garrett’s provocative findings (1998, Figs. 4.2, 5.2, 5.3, 5.4) appeared to demonstrate that in social-democratic and corporatist settings exposure to globalization pushes government spending upwards, while simultaneously enhancing these countries’ superior record of unemployment and economic growth. King et al., (2000) argued that they were able to provide an even stronger foundation for these conclusions by generating 1,000 sets of simulated coefficients and expected values for the scenarios contrasted in Garrett’s original study. Nevertheless, as shown by Garrett’s own data (1998, Figs. 3.10, 3.12), at least until very late in the period of the investigation his key scenarios actually had no empirical counterparts.

Chart 4 provides a graphical view of the limited empirical variability of the institutional configurations tapped by Garrett.²⁹ The X and Y axes measure his two dimensions of exposure to globalization – trade openness and restrictions on capital mobility. The bubbles that represent each country are proportional in size to Garrett’s index of “left-labor power”. It is evident that the 14 countries included in the study fall into a limited number of groups that exhaust only part of the available property space. In the upper half of the chart we find a social-democratic cluster with high levels of capital restrictions. The countries with fewer restrictions fall into two main groups.

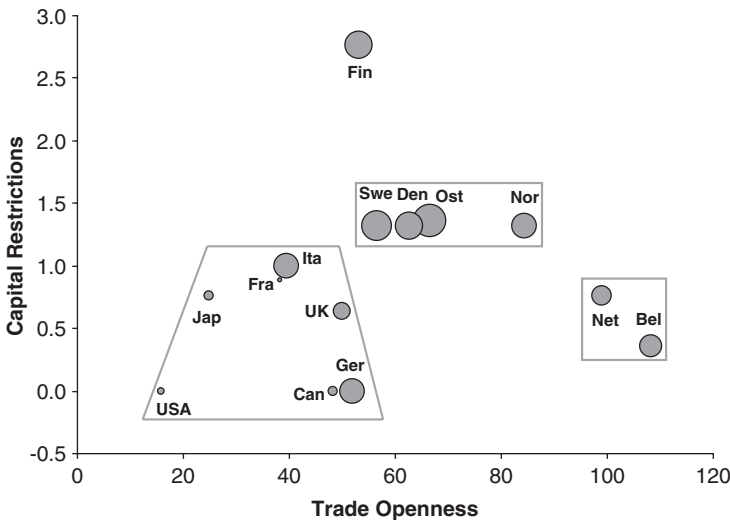


Chart 4. Garrett’s Indicators of Globalization (X and Y Axes) and Left-Labor Power (Bubbles).

Belgium and the Netherlands are small states highly involved in trade (cf. Katzenstein, 1985). The remaining seven countries are all large and relatively autarchic with few capital controls, although they exhibit diverse levels of labor strength. As a result of this clustering of Garrett's key variables it is evident for instance that no countries have either very high left power and unrestricted mobility, or low power and high trade openness. Despite this, Garrett calculated estimates of how the outcomes of interest would respond to high levels of globalization under both high and low left-labor power.³⁰

As King and Zeng (2002, p. 29) have argued in a different context, if “no evidence exists in our data with which to evaluate” a question, then “having time series–cross-sectional data with thousands of observations does not change this basic fact and will not make inferences like these any more secure”. This reinforces my earlier contention that investments in hi-tech statistical analysis are of limited value in fields like comparative political economy, where both the number of cases and their variability are severely restricted. Indeed, as Beck & Katz have wisely cautioned, “complicated methods often move us away from looking at and thinking about the data” (Beck & Katz, 1996, p. 31).

TESTING THE “REGIME” APPROACH

If the typical practitioner of pooling is guilty of closing his or her eyes to causal complexity, in *The Three Worlds of Welfare Capitalism* Gøsta Esping-Andersen (1990) took complexity as his essential starting-point. Unusually, Esping-Andersen combined and made explicit the desiderata posited by diverse traditions of comparative research: (1) recognizing that there may be striking causal discontinuities across different contexts; (2) informing hypotheses about relationships between variables by drawing on knowledge of cases; and (3) using quantitative indicators to systematically test propositions across the entire universe of cases. As this paper has tried to explain, while obviously consistent with the third of these goals MR is markedly inhospitable to the first two.

In his quantitative analysis, Esping-Andersen adopted a two-stage approach reminiscent of Hall and Franzese – first descriptive analysis and then MR. He developed indices of “universalism”, “decommodification” and “stratification” and used simple tables to show that his 18 OECD countries tend to fall into three distinct subgroups (Esping-Andersen, 1990, Tables 2.1, 3.3, 4.3). He then utilized MR to perform a causal analysis of cross-country variation in more than a dozen indicators, which were

regressed on political variables and in some cases control variables as well. However, Esping-Andersen's first technique (tabular analysis) was unnecessarily "soft", while the second (regression) is fundamentally in conflict with his analytical premises. There are better solutions, which exploit the rich data available on welfare states while respecting the theoretical assumption of causal complexity.

Esping-Andersen's tabular analysis relied heavily on his own judgment – both in the construction of indices and the identification of country clusters.³¹ No *systematic* test was carried out of whether his ensemble of indicators of welfare state regimes actually do "hang together"; and if they do, whether countries indeed cluster in three distinct subgroups on underlying policy dimensions. It would have been a logical step to subject these claims to techniques like factor analysis, cluster analysis, correspondence analysis or multidimensional scaling that seek to reveal underlying proximities between different variables or cases.

Demonstration of the existence of three policy regimes was of course only a preliminary to Esping-Andersen's search for empirical support for his causal arguments. Central here was his view that different welfare state regimes embody different socio-political forces and state traditions. Using MR, Esping-Andersen did his best to demonstrate that his preferred (political) explanations garnered stronger empirical support than rival (e.g. demographic) explanatory variables. These empirical results are of questionable value, being based on regressions with 5 or 6 explanatory variables and only 18 cases. The key difficulty, however, is that asking whether political effects "matter" after "controlling for" other causes is a different and more banal question than what actually interested Esping-Andersen. As stated in his own critique of the quantitative, cross-sectional research tradition, "The dominant correlational approach is ... marred by a frequent mismatch between theoretical intent and research practice" (Esping-Andersen, 1990, p. 106; see also Esping-Andersen, 1993).

The key causal argument of *The Three Worlds* is that *countries cluster on policy because they cluster on politics*. The regression approach, however, treats both policy and politics as continuous variables scattered across the whole spectrum of potential variation – not as a limited number of qualitatively different configurations with distinctive historical roots. In contrast to the causal thinking embodied in MR, Esping-Andersen would certainly *not* want to claim that, say, any discrete increment of Catholicism or absolutism ought to yield a discrete and uniform increment in the "corporativism" of pension programs. This is because *only* countries that are predominantly Catholic and/or have an absolutist past are expected to

exhibit the corporatist policy profile. By the same token, he would also not claim that the social policy of any given country may be understood precisely as the combined effect of Catholicism, absolutism and working class mobilization. (As in, “to make a loaf of bread combine 1 part yeast, 2 parts water and 10 parts flour ...”) On the contrary, a central purpose of his book was to demonstrate how the socialist, Catholic-Conservative and liberal political milieux have generated three different worlds of welfare. We may speculate that Esping-Andersen adopted MR out of deference to convention. He applied it as a blunt instrument for tapping gross differences between groups of countries, differences that arguably could have been more effectively conveyed by the use of tables and charts without the implication of constant linear effects across different contexts.³²

How might Esping-Andersen have exploited his quantitative data without falling back on the conventional statistical paradigm, which is so out of keeping with the spirit of his analysis and his critique of earlier work? Three early investigations offered innovative suggestions. [Ragin \(1994a\)](#) carried out an elaborate study of pension policy using seven different explanatory variables, by means of his own technique of qualitative comparative analysis (QCA). In the same volume [Kangas \(1994\)](#) compared the performance of QCA with cluster analysis and traditional regression techniques for testing a simplified political model of the quality of sickness insurance. A third study, by [Castles and Mitchell \(1992\)](#), used descriptive data to build an alternative typology of four overall worlds of welfare capitalism. Methodologically, while Castles and Mitchell refrained from going beyond the presentation of simplified tabular data, both Ragin and Kangas utilized cluster analysis to assign countries to regimes. But these creative efforts ran into serious difficulties. Kangas had trouble finding the Liberal countries and Ragin was placed in the awkward position of having to assign one third of his countries to a “spare” category, which automatically excluded them from his analysis. In performing cluster analysis of countries both authors were forcing them to fit into a single regime, thereby predetermining an issue in need of empirical exploration.³³

This issue has continued to bedevil subsequent research. A review by [Arts and Gelissen \(2002\)](#) concludes that Esping-Andersen’s typology has received only partial support from the empirical literature. According to these authors the typology is challenged because a significant number of countries lie between regimes. In their view, the imperfect fit between country cases and Esping-Andersen’s regimes indicates that more categories should be added to the typology. These conclusions reflect a common misunderstanding of the three worlds of welfare capitalism as referring literally to three discrete and mutually exclusive groupings of countries. However Esping-Andersen’s

core analytical concept was not “worlds” but “regimes”, that is to say *ideal-typical policy profiles*. As ideal-types they can be expected to resonate with the experience of some nations, but not to accurately describe all of them. On the contrary, hybrid cases are to be expected and the typology should help characterize and understand them more clearly. Finally, as already noted Esping-Andersen sees welfare regimes as reflecting three different political contexts. Hence the empirical usefulness of the regime typology should also be judged by whether countries’ placement with respect to regimes is paralleled by their political characteristics.

To summarize: (1) It is policy profiles and not necessarily countries that ought to follow a tripartite division; (2) The proximity or distance of a country’s policy profile from the three ideal-types should be matched by its political configuration; and (3) Policy regimes and their political underpinnings should together inform our understanding of individual countries. It follows that rather than seeking to assign countries to regimes, researchers should aspire to uncover underlying dimensions or profiles from cross-country correlations among policy indicators. Put differently, reducing a battery of *variables* to a few underlying dimensions is preferable to grouping *cases* into a few clusters. In light of this distinction it is not surprising that in Arts and Gelissen’s review of empirical tests of Esping-Andersen’s typology, the former methodology generated more supportive results than the latter.³⁴

Practically speaking, researchers interesting in uncovering policy regimes can choose from a variety of techniques, including factor analysis (Shalev, 1996) and its cousin, Principal Components Analysis (de Beer et al., 2001; Hicks & Kenworthy, 2003).³⁵ One of the attractive features of these methods of reducing data into a smaller number of dimensions is that they are not at all fazed by a multiplicity of variables. On the contrary, while the existence of a wealth of explanatory variables is the acknowledged bane of cross-national research, multiple indicators are actually desirable if the purpose is to more parsimoniously characterize the dependent variable.

What underlying dimensions would we expect to find if Esping-Andersen’s typology is correct? I believe that analytically his triplet of regimes rests on two dimensions of policy. One of them is a dichotomy that is unabashedly similar to Titmuss’ (1974) classic distinction between “residual” and “institutional” welfare state principles, often illustrated by contrasting the United States with Sweden. A second dimension, dubbed “corporativism” by Esping-Andersen, captures the fragmented, hierarchical and status-preserving measures pioneered by Catholic-Conservative welfare states, measures that were anathema to *both* socialist and bourgeois forces. It follows that if

Esping-Andersen is right about there being three ideal-typical worlds, we should be able to parsimoniously characterize the policies of actual welfare states in terms of these two dimensions.³⁶

Esping-Andersen’s original *The Three Worlds* volume identified several different loci of welfare state variation: social rights, social spending, the public/private division, and employment policy. The present reanalysis is based on 13 of Esping-Andersen’s policy indicators³⁷ and uses factor analysis to test whether the distribution of specific indicators follows the hypothesized two dimensions.³⁸ Factors are economical linear combinations of variables. They are generated in such a way that there is strong correlation between the variables with the highest “loadings” on a given factor, but minimal correlation between different factors (ideally they are completely uncorrelated or “orthogonal”).³⁹

The results of an unrotated principal component factor analysis are reported in Chart 5. The first two factors together account for the majority (nearly 60%) of the variance, good news for Esping-Andersen’s model. The first factor, which runs between the East and West of the chart, evidently captures the residual/institutional dimension. It exhibits high positive loadings on public employment, active labor market expenditure, benefit equality and social security spending; and strong negative loadings on poor relief and indicators of the scope of private health and pension provision. The

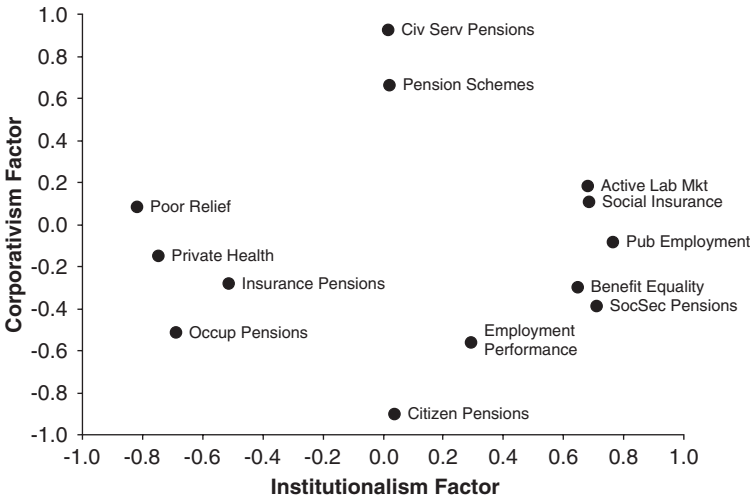


Chart 5. Two Factor Solution for Esping-Andersen Data.

second (North-South) factor signifies the corporatist dimension of policy. It has high positive loadings on the number of pension schemes and the prominence of civil service pensions, and a high negative loading on the role of “citizen pensions” (social security). The factors are not completely orthogonal, but the areas of overlap are intelligible. For instance, the results confirm that both the corporatist and institutional policy clusters are alienated from occupational pensions. They also imply that in the 1980s, when Esping-Andersen’s data were collected, employment performance (low unemployment and high job creation) was stronger in the institutional regime than in the residual or corporatist regimes.

We now evaluate Esping-Andersen’s political explanation for the origins of the three policy regimes. Chart 6 arrays the 18 nations in his study in accordance with their scores on our two factors. The evident linkage between policies and their political context generates an illuminating cross-national mapping. In particular, the findings support the clear distinction in Esping-Andersen’s (1990) book between the following three families of nations:

- *Socialist*: The Scandinavian social democracies, characterized by levels of working class mobilization almost without peer in other Western nations.
- *Catholic-Conservative*: Continental European nations – Italy, France, Belgium, Austria and Ireland – which share an absolutist past, relatively late-blooming democracy and a largely Catholic population.

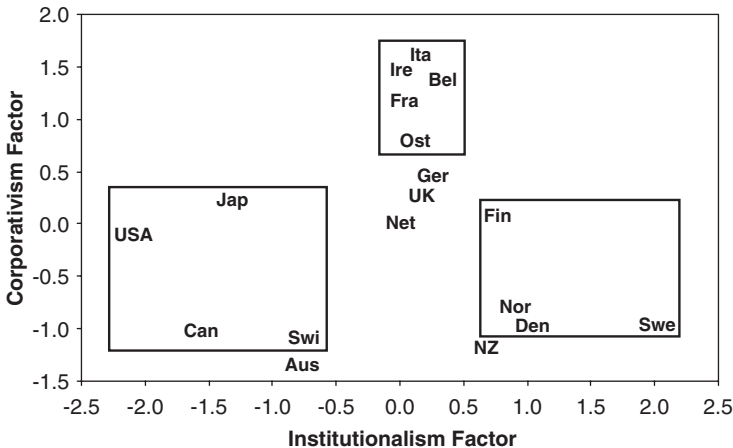


Chart 6. Social Policy Factor Scores.

- *Liberal*: The USA, Canada, Switzerland and Japan – in which working class mobilization is very weak and, in North America, the conservative heritage is absent.

The remaining five countries in Esping-Andersen's study are more difficult to classify. They have experienced moderate levels of working class mobilization but their state traditions are either close to the conservative group (Germany and the Netherlands), or were exposed in formative periods to liberal influences (the UK) or to the peculiar conditions of Antipodean settler societies (Australia and New Zealand).⁴⁰

The fit between the three political clusters and countries' placement on the two policy factors is substantial. The liberal states and Australia have the most negative institutionalism scores, while the Scandinavian states along with New Zealand have the highest positive scores. Most of the remaining countries are conservative states, and as expected they score indifferently on institutionalism but above average on the corporativism factor. Two mixed cases (Britain and the Netherlands) score close to zero on both factors, confirming their ambiguous status rather than making us wish they would go away.⁴¹

Our analysis largely supports Esping-Andersen's vision of three different policy constellations powered by three different constellations of political power. The key point is that this empirical support was garnered without the mismatch between ontology and methodology that is exemplified by the use of MR in *The Three Worlds*. Esping-Andersen's analytical reliance on ideal-types in the context of an ambitious program of comparative and historical research recalls the classic sociological tradition, one which continues to inspire many comparativists. His goal of subjecting the theory of welfare state regimes to systematic empirical test was also admirable, but MR was ill-suited to this task. I have tried to show that methodological alternatives are available which do not require sacrificing either quantification or the ambition of supporting causal claims through empirical generalization.

CONCLUSION

Despite considerable methodological debate and innovation among comparativists in recent years, MR remains by far the predominant mode of numerical data analysis and most of its critics see qualitative analysis (whether formal or not) as the only real alternative. This paper seeks to promote a third way. I recognize that Charles Ragin's innovations, QCA

and more recently “fuzzy-set” analysis (Ragin, 1987, 2000), point to another strategic alternative. Ragin’s techniques constitute a synthesis of the qualitative and quantitative traditions aimed at explicitly testing the kind of “causal pathways” arguments typical of classical comparative-historical research in the genre of Weber, Moore, Rokkan and Skocpol. The desire to systematically evaluate the evidence for such arguments is not new (Somers, 1971). But Ragin (1987) is the first to have offered formal procedures for parsimoniously identifying the regularities that underlie a series of case configurations.

Ragin’s methods are not “qualitative” in the sense of relying on the interpretive skills of analysts wading knee-deep in thick description. If anything, as Griffin and Ragin (1994, p. 10) have insisted, QCA is more like MR: both apply rules that are independent of the researcher, and both treat cases as “discrete, multiple instances of more general phenomena”. While controversial,⁴² in principle Ragin’s methods have great advantages because of their fidelity to principles of case-oriented analysis. One feature, which is especially valuable in the context of small-n macro-comparisons, but lacking in MR, is visibility of and dialog with the cases. However, the advantages of Ragin’s techniques are not exclusive to his methods. My reanalysis of diverse MR-based studies in this paper poses alternatives to both QCA and MR. In closing, I incorporate these suggestions into a summary statement of the major options (other than Ragin’s methods) open to quantitative researchers who are troubled by the limitations of MR.

1. *Refinement.* This is the optimistic approach best represented in the present survey by Bruce Western’s variant of pooled regression. However, the discovery of a serious limitation of Western’s method heightens our pessimism concerning the payoffs from technical refinement. Western was unable to resolve the problem of simultaneously combining and separating cross-country and over-time effects. This is only one issue in MR analysis for which political scientists have sought inspiration from their technically more advanced counterparts in economics and statistics. In this connection it is sobering that G.S. Maddala, one of the most respected figures in the econometric world, considers its achievements both modest and contested. Moreover, he believes that leading political methodologists have mistakenly or misguidedly emulated shallow econometric fads (Maddala, 1998). Sadly, Maddala’s criticisms and cautions appear to have fallen on deaf ears.⁴³ More encouraging is the emerging trend, noted earlier, of efforts to find original econometric solutions to some of the lacunae of MR highlighted in this paper. However it is too

early to predict the fate of these new methods. They are as likely to spark new rounds of technical debate or simply be ignored as to triumph over researchers' customary methodological conservatism.

2. *Triangulation*. This means combining MR with other types of analysis – quantitative, qualitative or both. Hall and Franzese adopted this approach to strengthen their empirical case by citing the convergent findings produced by different ways of researching the same topic. Alternatively, the complementarity of different approaches may rest on the distinctive contributions made by each one of them. This is the strategy underpinning Esping-Andersen's work on welfare states, and several ambitious comparative and historical studies by John Stephens, Evelyn Huber & their collaborators (Rueschemeyer et al., 1992; Huber & Stephens, 2001; see also Huber, Ragin, & Stephens, 1991; Rueschemeyer & Stephens, 1997). They have proposed that comparative research be based on dialog between broad-spectrum quantitative comparisons and historically oriented country studies (see also Esping-Andersen, 1993). The results of MR should be confronted by both theory and knowledge of cases, and if causal anomalies arise they should be put to the test of historical process-tracing across multiple countries.

This approach is attractive but also very demanding; it is virtually impossible without long-term collaborative research. In practice, when triangulation does occur it is usually more modest than in the hands of Stephens and his collaborators. Occasionally, researchers employ multiple statistical techniques to analyze the same data or problem, looking for convergent results (e.g. the use of both MR and QCA by Kangas, 1994; Ebbinghaus & Visser, 1999). In addition, some book-length studies have utilized both case-studies and pooled regressions, using the qualitative materials either to illustrate their argument (e.g. Boix, 1998) or as a genuine complement to statistical findings (e.g. Swank, 2002).⁴⁴ This kind of hybrid analysis is a welcome development, but the insularity of different methodological traditions and the difficulty of publishing multimethod articles in journal format both limit its likely spread.

3. *Substitution*. The present paper has promoted the use of alternative methods of quantitative analysis as another strategy for dealing with the problems of MR. The second and third sections presented tables or tree diagrams in which countries are clearly identified.⁴⁵ It was shown that these simple techniques overcome some of the most unattractive limitations of MR while incorporating key elements of the case-oriented approach. They are able to plainly convey complex analytical ideas like elective affinities and causal hierarchies. They also draw attention to cases

deserving of additional, more focused comparative scrutiny, which is a blind spot of most other methods. I have suggested as well that, provided they fit researchers' theoretical assumptions, there is no reason why inductive multivariate statistical methods should not be exploited by comparativists. The utility of factor analysis in clarifying the evidence for Esping-Andersen's approach to welfare state diversity was the illustration offered here,⁴⁶ but many other methods of exposing latent variables are available. Such methods hold the delicious promise of turning the traditional handicap of more indicators than cases from a burden into an asset. Of course, generating better measures of the phenomena of interest cannot resolve the difficulties of testing causal explanations in cross-national research. It has been argued here that data analysis aimed at theory testing and theory building should strive to reveal how the cases are located in relation to each other as well as to cause and effect variables.

ACKNOWLEDGMENTS

Participants at several workshops and conferences where this paper was presented were kind enough to offer comments and advice. In addition I wish to acknowledge valuable input from Neal Beck, Frank Castles, David Freedman, Peter Hall, Robert Franzese, Orit Kedar, Bernhard Kittel, Walter Korpi, Noah Lewin-Epstein, Hadas Mandel, Jonathon Moses, Herbert Obinger, Meir Shabat, Aage Sorensen, David Soskice, John Stephens, Uwe Wagschal and Bruce Western.

NOTES

1. In contrast, interest in formal methods tailored to small-n research is relatively strong in Europe, with an extensive website devoted to the topic (<http://www.compass.org>).

2. It is, of course, debatable just how bounded the research universe is or should be. Conventionally, comparative policy studies focus on the approximately 18 rich, capitalist countries with longstanding democratic polities and non-trivial populations. Such conventions may be theoretically arbitrary and should always be open to challenge. Many studies have incorporated Greece, Spain and Portugal after democratization (and more practically, after their inclusion in OECD databases). Other candidates for inclusion in studies of what have until now been known as "the Western nations" might be found in the former Soviet bloc states, Latin America and East Asia. There are good arguments both for and against expanding the universe of comparative studies. For instance, compare Geddes (1990) and Boyer (1997).

3. Even the well-known injunction of Przeworski and Teune (1970) that comparativists should strive to turn the proper names of countries into the abstract names of variables did not entirely contradict this view. It should be remembered that Przeworski and Teune were railing against the dominance of comparative politics by “area studies” specialists and urging their colleagues to avoid particularizing arguments that could easily strait-jacket both theory and comparison. Many contemporary advocates of case-oriented analysis (including Ragin) would have no quarrel with this assessment.

4. The criticism here is not the standard one that quantification over-simplifies complex reality. There is always a trade-off between accuracy and parsimony in social research, whether analysis uses quantitative measures or narrative representations. The point is that the use of MR encourages what may well be a mistaken belief that our measures are precise and continuous.

5. An exception is Amenta and Poulson’s (1996) use of MR and QCA in a comparative study of the American states. This exception proves the rule, however, since the measurement of such concepts as “administrative strength” was possible only because this research compared sub-national units of a uniform national entity.

6. More recently, Lieberman and Lynn (2002) have offered a more fundamental critique of the quasi-experimental epistemology prevalent in sociology and similar disciplines.

7. Abbot has offered an elegant formulation of this problem. Variable-oriented approaches “seek to understand the social process by developing linear transformations from a high-dimensional space (of ‘main effects’ and occasionally of interactions between them) into a single dimension (the dependent variable) ... Now this strategy ... is useful only if the data space is more or less uniformly filled” (Abbott, 1997, p. 86).

8. I excluded picayune Iceland with only 80,000 potential union members. I also replaced Rothstein’s left party representation indicator borrowed from Wilensky (1981) and based on the entire 1919–1979 period which includes disruptions and discontinuities during the interwar years. Since the unionization data reveal that cross-national differentials stabilized after about 1965, I treat the first two postwar decades as the politically formative period. Figures for average left cabinet strength in this period were taken from the dataset assembled by Korpi and Shalev (1980). It turns out that these modifications strengthen the effect of the Ghent variable.

9. Potential membership was dichotomized after exploratory charts revealed that it had an evident threshold effect on unionization. With the exceptions of only Switzerland and the Netherlands, all small countries (no more than 5 million potential members) had more than 50% density, while all the large countries (10 million and up) scored less than 50%. Within these two categories no relationship was discernible between the two variables.

Left strength was grouped into four categories that reflect breaks in its distribution. “None” were cases with zero or trivial (up to Japan’s 4%) left party representation in cabinet; “weak” 7–15%; “medium” 22–29% plus an intermediate case (the UK) with 36%; “strong” 45% or more.

10. On the other hand, left strength discriminates only weakly between the unionization rates of small countries, and not at all between the large ones (except perhaps for the British case).

11. It should be pointed out however that although only careful comparative historical research can speak to this type of causal question, as a result of theoretical, evidentiary and interpretive differences there is no guarantee that a consensual account will emerge. On the contrary, a sizable literature relevant to the role of the Ghent system has failed to arrive at clear-cut conclusions. In addition to Rothstein's article, see Hancke (1993), Scruggs (2002), Oskarsson (2003) and Swenson (2002).

12. The significance of these kinds of anomalies for scientific progress has been strongly argued by Rogowski (1995).

13. Visser (1992) has suggested that most of the vast difference between Belgian and Dutch unionization can be attributed to the fact that Dutch unions have no presence in the workplace. The origins of Norway's laggard status are less clear, but they might be traceable to the Norwegian union movement's lesser effectiveness in some of the sectors that grew from the 1960s, when Norway's density plateaued while Sweden's entered a long period of growth. Data collected by D'Agostino (1992) reveal substantial gaps in union density favoring Sweden in the following categories: women, private sector trade and services, and white-collar workers.

14. See http://www-personal.umich.edu/~franzese/h&f_data.TXT

15. The assumption of causal stability over time can be relaxed, but as in Hall and Franzese's study it typically is not. Although Hall and Franzese tested for effects of different data periodicities (annual, decadal or full-period), they did not examine the consistency of their model across sub-periods.

16. Except at the intermediate level of coordination (0.5), where there is only a small difference in CBI between Belgium and the Netherlands. Since the Hall–Franzese model in any case makes no specific prediction for this configuration I do not include it in Table 1.

17. Hall and Franzese included dummy variables for each decade or year in their pooled regressions, but they were not interacted with any of the causal variables.

18. Hall and Franzese's simulation estimated 9.7 percentage points more unemployment at the highest than the lowest levels of CBI in decentralized systems, whereas the simulated gap between the actually existing poles of CBI is only 2.4 points.

19. Goldthorpe recommends even more strongly that researchers widen the "geographical and sociocultural range" of their research. In this matter, however, it cannot be said (as it can of pooling) that the recommended solution is a popular one. As Goldthorpe concedes, data quality and availability are limited outside of the bloc – the OECD countries – which interests his intended audience (and mine). Moreover it is widely understood that what might be called the "specification costs" of going beyond the OECD (additional casual factors and alternative causal paths) usually outweigh the potential benefits. Even in a theoretically developed field (the economics of growth) where it was possible to gather comparable data for a stunning 119 countries, Levine and Renelt (1992) found themselves hopelessly unable to use cross-national regressions to adjudicate between rival theories.

20. In political science, where pooling has been most popular, foundational treatments are Stimson (1985), Sayers (1989) and Hicks (1994).

21. In Kittel and Winner's (2005, p. 8) pithy summary, "practically all published contributions to comparative political economy using panel data assume poolability by fiat".

22. A compromise that is more sensitive to context but less exhaustive of degrees of freedom, is to permit both intercept and slope parameters to vary across *groups* of nations or years. For a rare example see O'Connell (1994).

23. Western's 1998 article is the published version of a paper dated December 1996 which was circulated electronically (Western, 1996). In the final version a partly different empirical example was substituted for the one in the preprint version (economic growth became the dependent variable instead of unemployment). I refer here to the findings reported in the 1996 version since they highlight a problem, which I believe to be endemic to the technique that Western proposed.

24. "Unregulated" labor markets are those in which no more than half of the workforce was covered by collective bargaining. Classification of the other countries was based on Western's dichotomous measure of corporatism. I adopted Western's classification of Switzerland as corporatist even though it had less than 50% collective bargaining coverage.

25. For example, if the time-series coefficients for left cabinet strength are regressed cross-nationally on collective bargaining coverage, the resulting coefficient is 1.00 ($t = 3.4$) for all countries but only 0.59 ($t = 1.5$, non-significant) without Japan and Finland.

26. Western (1996, p. 26) indeed noted that the left government variable for Japan was constant and counseled against "substantive interpretation" of the Japanese result. However the statistical generalization yielded by the cross-sectional level of his hierarchical model was clearly based in part on the Japanese case.

27. The debate took place in a special issue of *International Organization*. For a judicious summary, see the contribution by King (2001).

28. For additional wider-ranging critiques of Garrett's study, see Hay (2000) and Moses (2001).

29. Chart 4 is based on averages for the full period of Garrett's investigation (1966–1990) which I calculated using the dataset on his Yale University website (<http://pantheon.yale.edu/~gmg8>) in August 2000.

30. In a private communication dated March 7, 2001, Garrett concurred that with one temporary and partial exception no country in his dataset with a strong left exhibited weak capital controls, but he argued that out-of-sample experience in the 1990s subsequently vindicated his predictions.

31. Recent research has sought to replicate and/or update Esping-Andersen's de-commodification scores. Lyle Scruggs is highly critical of Esping-Andersen's methodology (see his "Comparative Welfare State Entitlements" website at <http://sp.uconn.edu/~scruggs/wp.htm> and Scruggs and Allen (2006)), while Bambra (2004) reports similar results to Esping-Andersen using updated sources.

32. In his more recent work Esping-Andersen (1999) adopted a different variant of MR, multinomial logistic regression. In keeping with the spirit of the regime approach, this technique has the advantage of permitting explanatory weights to vary across different categories of the dependent variable. But in the context of cross-national research of this type, the category-specific coefficients must be estimated on ludicrously small numbers of cases.

33. Both of the standard approaches to clustering – hierarchical and *k*-means – allocate cases to mutually exclusive clusters, although they provide information on how well each case fits its group.

34. For an exception published after Arts and Gelissen's survey see [Powell and Barrientos \(2004\)](#).

35. In addition to the techniques mentioned, other methods of revealing underlying "dimensions" are MDS (multidimensional scaling) and CA (correspondence analysis). These methods are appropriate to ordinal or even nominal data and do not assume linear relationships among variables. Another flexible option, utilized by [de Beer, Vrooman, and Wildeboer Schut \(2001\)](#), is the non-linear version of Principal Components Analysis known in SPSS as PRINCALS. Since the results generated by factor analysis in my original study ([Shalev, 1996](#)) are replicated using other methods, they remain the basis for the findings reported here.

36. [Hicks and Kenworthy \(2003\)](#) also advocate a dimensional approach to verifying Esping-Andersen's typology. However, these authors seem to interpret their finding that welfare state indicators reduce to two dimensions as evidence against the existence of three regimes. In contrast, I argue that if Esping-Andersen is correct then policies (again – *not* countries) should follow two underlying continua which provide the coordinates of the three regimes.

37. In view of objections raised by [Castles and Mitchell \(1992\)](#) concerning his coding of Australia and New Zealand, I did not include two of Esping-Andersen's key indicators – "decommodification" and "universalism" ([Esping-Andersen, 1990](#), Tables 2.2, 3.1). The 13 indicators summarized in Chart 4 were obtained as follows; references are to [Esping-Andersen \(1990\)](#): social insurance spending (Table 5.1, source data from the author); number of pension schemes ("Corporatism" in Table 3.1), Civil Servants' pensions ("Etatism" in Table 3.1), benefit equality (Table 3.1); "poor relief" (Table 3.1); the public-private division in health (Table 3.1) and pensions (Table 4.3); "full-employment performance" (Table 5.9, data from the author). Active manpower program expenditures relative to GDP (c. 1975) and public employment as a percentage of total employment (in 1980) are mentioned in [Esping-Andersen \(1990\)](#) and analyzed in [Esping-Andersen \(1985\)](#), but the source data were obtained directly from the author.

38. The findings presented below were originally reported in the introduction to [Shalev \(1996\)](#).

39. Thus the researcher hopes that each item will load high on only one of the factors. The procedure known as factor "rotation" is designed to encourage this to happen, but I opted here for the more pristine test of an unrotated analysis.

40. On the complexity and importance of state traditions as a causal variable in comparative research, see [Crouch \(1993\)](#).

41. The contradictions of the British welfare state are well known, and if anything they are exemplified by the contrasting experiments launched by Thatcher and Blair. On the mixed Dutch case, see [Wildeboer Schut, Vrooman, and de Beer \(2001\)](#).

42. QCA has been vociferously criticized, particularly for its dichotomous measurement of variables and abandonment of probabilistic generalizations in favor of deterministic ones (see especially [Liebersson, 1994, 1991](#); [Goldthorpe, 1997](#)). Ragin's "fuzzy logic" technique at least partially answers these criticisms.

43. In quest of evidence for political methodologists' inattention to critiques of pooling, I used the Social Sciences Citation Index to search for articles that cited [Maddala \(1998\)](#). As of July 1, 2005, there were only five citations, two of them authored by political methodologists. In contrast, another article by [Maddala](#)

(on unit roots and cointegration) published the same year has been cited more than 100 times.

44. In an intriguing recent contribution, Gordon and Smith (2004) offer a method for introducing qualitative findings into causal statistical models (which however has already given rise to debate; see *Political Analysis*, Vol. 13, No. 3).

45. For an independent application of these techniques, see Marks and Wilson (2000, pp. 445, 450).

46. See also Leertouwer (2002), who used factor analysis to uncover the latent dimensions of corporatism and central bank independence by analyzing a wide range of empirical indicators proposed by previous researchers.

REFERENCES

- Abbott, A. (1997). On the concept of turning point. *Comparative Social Research*, 16, 85–105.
- Abbott, A. (1998). The causal devolution. *Sociological Methods and Research*, 27(2), 148–181.
- Achen, C. H. (2000). Why lagged dependent variables can suppress the explanatory power of other independent variables. Paper presented at the annual meeting of the political methodology section of the American Political Science Association, UCLA, July 20–22, <http://web.polmeth.ufl.edu/papers/00/achen00.zip>
- Alvarez, R. M., Garrett, G., & Lange, P. (1991). Government partisanship, labor organization, and macroeconomic performance. *American Political Science Review*, 85(2), 539–556.
- Amenta, E. (1993). The state of the art in welfare state research on social spending efforts in capitalist democracies since 1960. *American Journal of Sociology*, 99(3), 750–763.
- Amenta, E., & Poulsen, J. D. (1996). Social politics in context: The institutional politics theory and social spending at the end of the New Deal. *Social Forces*, 75(1), 33–59.
- Arts, W., & Gelissen, J. (2002). Three worlds of welfare capitalism or more? A state-of-the-art report. *Journal of European Social Policy*, 12(2), 137–158.
- Bambra, C. (2004). Weathering the storm: Convergence, divergence and the robustness of the ‘worlds of welfare’. *The Social Policy Journal*, 3(3), 3–23.
- Beck, N., & Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review*, 89(3), 634–647.
- Beck, N., & Katz, J. N. (1996). Nuisance vs. substance: Specifying and estimating time-series cross-section models. *Political Analysis*, 6, 1–36.
- Beck, N., & Katz, J. N. (2003). Random coefficient models for time-series–cross-section data. Paper presented at the 2nd ECPR conference, Marburg (Germany), 18–21 September, <http://www.essex.ac.uk/ecpr/events/generalconference/marburg/papers/6/2/Katz.pdf>
- Beck, N., Katz, J. N., Alvarez, R. M., Garrett, G., & Lange, P. (1993). Government partisanship, labor organization, and macroeconomic performance – a corrigendum. *American Political Science Review*, 87(4), 945–954.
- Berg-Schlosser, D. (2002). Macro-quantitative vs. macro-qualitative methods in the social sciences – testing empirical theories of democracy. Paper presented at the XV world congress of the International Sociological Association, Brisbane, July 7–3.
- Berg-Schlosser, D., & De Meur, G. (1997). Reduction of complexity for a small-N analysis: A stepwise multi-methodological approach. *Comparative Social Research*, 16, 133–162.
- Boix, C. (1998). *Political parties, growth and equality: Conservative and social democratic economic strategies in the world economy*. Cambridge: Cambridge University Press.

- Boyer, R. (1997). Comparing Germany and Japan: Methodological issues, main findings from the regulation approach, and an agenda for further research. Paper presented at the conference on Germany and Japan: The Future of Nationally Embedded Capitalism in a Global Economy, University of Washington, Seattle, April 10–12.
- Braumoeller, B. F. (2003). Causal complexity and the study of politics. *Political Analysis*, 11(3), 209–233.
- Braumoeller, B. F., & Goertz, G. (2000). The methodology of necessary conditions. *American Journal of Political Science*, 44(4), 844–858.
- Cameron, D. R. (1984). Social democracy, corporatism, labour quiescence and the representation of economic interest in advanced capitalist society. In: J. H. Goldthorpe (Ed.), *Order and conflict in contemporary capitalism* (pp. 143–178). Oxford: Clarendon Press.
- Castles, F. G. (1993). *Families of nations: Patterns of public policy in western democracies*. Brookfield, VT: Dartmouth.
- Castles, F. G., & Mitchell, D. (1992). Identifying welfare state regimes: The links between politics, instruments, and outcomes. *Governance*, 5(1), 1–26.
- Crouch, C. (1993). *Industrial relations and European state traditions*. Oxford: Clarendon Press.
- D'Agostino, H. (1992). *Why do workers join unions? A comparison of Sweden and OECD countries*. Stockholm: SOFI (Swedish Institute for Social Research).
- de Beer, P., Vrooman, C., & Wildeboer Schut, J. M. (2001). *Measuring welfare state performance: Three or two worlds of welfare capitalism?* Luxembourg Income Study Working Paper no. 276, May.
- Ebbinghaus, B., & Visser, J. (1999). When institutions matter: Union growth and decline in Western Europe, 1950–1995. *European Sociological Review*, 15(2), 135–158.
- Esping-Andersen, G. (1985). Power and distributional regimes. *Politics and Society*, 14(2), 223–256.
- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge: Polity Press.
- Esping-Andersen, G. (1993). The comparative macro-sociology of welfare states. In: L. Moreno (Ed.), *Social exchange and welfare development* (pp. 123–136). Madrid: CSIS.
- Esping-Andersen, G. (1999). *Social foundations of postindustrial economies*. Oxford: Oxford University Press.
- Esping-Andersen, G., & Korpi, W. (1984). Social policy as class politics in post-war capitalism: Scandinavia, Austria, and Germany. In: J. H. Goldthorpe (Ed.), *Order and conflict in contemporary capitalism* (pp. 179–208). Oxford: Oxford University Press.
- Firebaugh, G. (1980). Cross-national versus historical regression models: Conditions of equivalence in comparative research. *Comparative Social Research*, 3, 333–344.
- Franzese, R. J. (2001). *Macroeconomic policies of developed democracies*. New York: Cambridge University Press.
- Freedman, D. (1985). Statistics and the scientific method. In: W. M. Mason & S. E. Fienberg (Eds), *Cohort analysis in social research: Beyond the identification problem* (pp. 345–390). New York: Verlag.
- Freedman, D. (1987). As others see us: A case study in path analysis (with discussion). *Journal of Educational Statistics*, 12, 101–223.
- Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology*, 21, 291–313.
- Garrett, G. (1998). *Partisan politics in the global economy*. Cambridge: Cambridge University Press.

- Garrett, G., & Lange, P. (1989). Government partisanship and economic performance – when and how does who governs matter. *Journal of Politics*, 51(3), 676–693.
- Garrett, G., & Mitchell, D. (2001). Globalization, government spending and taxation in the OECD. *European Journal of Political Research*, 39(2), 145–177.
- Geddes, B. (1990). How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis*, 2, 131–150.
- Giroi, F., & King, G. (2001). *Time series cross-sectional analyses with different explanatory variables in each cross-section. The Global Burden of Disease 2000 In Aging Populations*. Research Paper no. 10, Harvard University, July, [http://www.hsph.harvard.edu/burdenofdisease/publications/papers/Time Series Cross-Sectional Analyses.pdf](http://www.hsph.harvard.edu/burdenofdisease/publications/papers/Time%20Series%20Cross-Sectional%20Analyses.pdf)
- Goldthorpe, J. H. (1997). Current issues in comparative macrosociology: A debate on methodological issues. *Comparative Social Research*, 16, 1–26.
- Goodman, L. A. (1981). Criteria for determining whether certain categories in a cross-classification table should be combined, with special reference to occupational categories in an occupational mobility table. *American Journal of Sociology*, 87(3), 612–650.
- Gordon, S. C., & Smith, A. (2004). Quantitative leverage through qualitative knowledge: Augmenting the statistical analysis of complex causes. *Political Analysis*, 12(3), 233–255.
- Griffin, L., Barnhouse Walters, P., O'Connell, P., & Moor, E. (1986). Methodological innovations in the analysis of welfare-state development: Pooling cross sections and time series. In: N. Furniss (Ed.), *The futures for the welfare state* (pp. 101–138). Bloomington: Indiana University Press.
- Griffin, L. J., O'Connell, P. J., & McCammon, H. J. (1989). National variation in the context of struggle: Post-war class conflict and market distribution in the capitalist democracies. *Canadian Review of Sociology and Anthropology*, 26(1), 37–68.
- Griffin, L., & Ragin, C. C. (1994). Some observations on formal methods of qualitative analysis. *Sociological Methods & Research*, 23(1), 4–21.
- Hall, P. A. (2003). Aligning ontology and methodology in comparative research. In: J. Mahoney & D. Rueschmeyer (Eds), *Comparative historical research in the social sciences*, (pp. 373–404). New York: Cambridge University Press.
- Hall, P. A., & Franzese, R. J., Jr. (1998). Mixed signals: Central bank independence, coordinated wage bargaining, and European monetary union. *International Organization*, 52(3), 505–535.
- Hancke, B. (1993). Trade union membership in Europe, 1960–1990 – rediscovering local unions. *British Journal of Industrial Relations*, 31(4), 593–613.
- Hay, C. (2000). Globalization, social democracy and the persistence of partisan politics: A commentary on Garrett. *Review of International Political Economy*, 7(1), 138–152.
- Hibbs, D. A., Jr. (1976). Industrial conflict in advanced industrial societies. *American Political Science Review*, 70(4), 1033–1058.
- Hibbs, D. A., Jr. (1978). On the political economy of long-run trends in strike activity. *British Journal of Political Science*, 8(2), 153–175.
- Hicks, A. (1994). Introduction to pooling. In: T. Janoski & A. M. Hicks (Eds), *The comparative political economy of the welfare state* (pp. 169–188). Cambridge: Cambridge University Press.
- Hicks, A., & Kenworthy, L. (2003). Varieties of welfare capitalism. *Socio-Economic Review*, 1(1), 27–61.

- Huber, E., Ragin, C., & Stephens, J. D. (1991). Quantitative studies of variation among welfare states: Towards a resolution of the controversy. Paper presented at the ISA conference on comparative studies of welfare state development, Helsinki.
- Huber, E., Ragin, C., & Stephens, J. D. (1993). Social democracy, Christian democracy, constitutional structure, and the welfare state. *American Journal of Sociology*, 99(3), 711–749.
- Huber, E., & Stephens, J. D. (2001). *Development and crisis of the welfare state: Parties and policies in global markets*. Chicago: The University of Chicago Press.
- Iversen, T. (1999). *Contested economic institutions: The politics of macroeconomics and wage bargaining in advanced democracies*. New York: Cambridge University Press.
- Iversen, T., Pontusson, J., & Soskice, D. W. (1999). *Unions, employers, and central banks: Macroeconomic coordination and institutional change in social market economies*. New York: Cambridge University Press.
- Janoski, T., & Hicks, A. M. (1994). *The comparative political economy of the welfare state*. Cambridge: Cambridge University Press.
- Kangas, O. (1994). The politics of social security: On regressions, qualitative comparisons and cluster analysis. In: T. Janoski & A. M. Hicks (Eds), *The comparative political economy of the welfare state* (pp. 346–364). Cambridge: Cambridge University Press.
- Katzenstein, P. (1985). *Small states in world markets: Industrial policy in Europe*. Ithaca, NY: Cornell University Press.
- Kenworthy, L. (2001). Wage-setting measures – a survey and assessment. *World Politics*, 54(1), 57–98.
- King, G. (2001). Proper nouns and methodological propriety: Pooling dyads in international relations data. *International Organization*, 55(2), 497–507 U7.
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 44(2), 347–361.
- King, G., & Zeng, L. (2002). *When can history be our guide? The pitfalls of counterfactual inference*. Unpublished paper, Harvard University Department of Government, May 17, <http://gking.harvard.edu/files/counterf.pdf>
- Kittel, B. (1999). Sense and sensitivity in pooled analysis of political data. *European Journal of Political Research*, 35(2), 225–253.
- Kittel, B., & Winner, H. (2005). How reliable is pooled analysis in political economy? The globalization-welfare state nexus revisited. *European Journal of Political Research*, 44(1), 1–25.
- Korpi, W. (1983). *The democratic class struggle*. London: Routledge and Kegan Paul.
- Korpi, W. (1991). Political and economic explanations for unemployment: A cross-national and long-term analysis. *British Journal of Political Science*, 21(3), 315–348.
- Korpi, W., & Shalev, M. (1980). Strikes, power and politics in the western nations, 1900–1976. *Political Power and Social Theory*, 1, 301–334.
- Lange, P., & Garrett, G. (1985). The politics of growth: Strategic interaction and economic performance in the advanced industrial democracies, 1974–1980. *Journal of Politics*, 47, 792–827.
- Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73(1), 31–44.

- Leertouwer, E. (2002). *Measurement issues in political economy*. Groningen: University of Groningen SOM, <http://www.ub.rug.nl/eldoc/dis/eco/e.c.leertouwer>
- Levine, R., & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 82(4), 942–963.
- Lieberson, S. (1985). *Making it count*. Berkeley, CA: University of California Press.
- Lieberson, S. (1991). Small N's and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases. *Social Forces*, 70(2), 307–320.
- Lieberson, S. (1994). More on the uneasy case for using Mill-type methods in small-N comparative studies. *Social Forces*, 72(4), 1225–1237.
- Lieberson, S., & Lynn, F. B. (2002). Barking up the wrong branch: Scientific alternatives to the current model of sociological science. *Annual Review of Sociology*, 28, 1–19.
- Maddala, G. S. (1998). Recent developments in dynamic econometric modelling: A personal viewpoint. *Political Analysis*, 7, 59–87.
- Marks, G., & Wilson, C. J. (2000). The past in the present: A cleavage theory of party response to European integration. *British Journal of Political Science*, 30(3), 433–459.
- Martin, A. (1973). *The politics of economic policy in the United States: A tentative view from a comparative perspective*. Beverly Hills: Sage (Sage Professional Papers in Comparative Politics, No. 01-040).
- Moses, J. W. (2001). A methodological critique of statistical studies of globalization. Paper presented to the Max Planck Institute for the study of societies, Cologne, 7 June.
- O'Connell, P. (1994). National variation in the fortunes of labor: A pooled and cross-sectional analysis of the impact of economic crisis in the advanced capitalist nations. In: T. Janoski & A. M. Hicks (Eds), *The comparative political economy of the welfare state* (pp. 218–242). Cambridge: Cambridge University Press.
- Oskarsson, S. (2003). Institutional explanations of union strength: An assessment. *Politics & Society*, 31(4), 609–635.
- Paldam, M., & Pedersen, P. J. (1982). The macroeconomic strike model: A study of 17 countries, 1948–1975. *Industrial and Labor Relations Review*, 35(4), 504–521.
- Plumper, T., Troeger, V. E., & Manow, P. (2005). Panel data analysis in comparative politics: Linking method to theory. *European Journal of Political Research*, 44(2), 327–354.
- Podesta, F. (2003). Econometric solutions vs. substantive results: A crucial tradeoff in the time-series–cross-section analysis. Paper presented at the 2nd ECPR conference, Marburg (Germany), 18–21 September, <http://econpapers.repec.org/paper/eseiserwp/2003-34.htm>
- Powell, M., & Barrientos, A. (2004). Welfare regimes and the welfare mix. *European Journal of Political Research*, 43(1), 83–105.
- Przeworski, A., & Tuene, H. (1970). *The logic of comparative social inquiry*. New York: Wiley.
- Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley, CA: University of California Press.
- Ragin, C. C. (1994a). A qualitative comparative analysis of pension systems. In: T. Janoski & A. M. Hicks (Eds), *The comparative political economy of the welfare state* (pp. 320–345). Cambridge: Cambridge University Press.
- Ragin, C. C. (1994b). *Constructing social research: The unity and diversity of method*. Thousand Oaks, CA: Pine Forge Press.
- Ragin, C. C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press.
- Rogowski, R. (1995). The role of theory and anomaly in social-scientific inference. *American Political Science Review*, 89(2), 467–470.

- Rothstein, B. (1990). Labour market institutions and working class strength. In: S. Steinmo, K. Thelen & F. Longstreth (Eds), *Structuring politics: Historical institutionalism in comparative analysis* (pp. 33–56). Cambridge: Cambridge University Press.
- Rueschemeyer, D., Huber-Stephens, E., & Stephens, J. D. (1992). *Capitalist development and democracy*. Cambridge: Polity Press.
- Rueschemeyer, D., & Stephens, J. D. (1997). Comparing historical sequences – a powerful tool for causal analysis. In: L. Mjøset & F. Engelstadt (Eds), *Methodological issues in comparative social science* (pp. 33–56). Greenwich, CT: JAI Press.
- Sayrs, L. (1989). *Pooled time series analysis*. London: Sage (Quantitative Applications in the Social Sciences No. 70).
- Scruggs, L. (2002). The Ghent system and union membership in Europe, 1970–1996. *Political Research Quarterly*, 55(2), 275–297.
- Scruggs, L., & Allen, J. (2006). Welfare-state decommodification in 18 OECD countries: A replication and revision. *Journal of European Social Policy*, 16(1), 55–72.
- Shalev, M. (1979a). *The strike: Theoretical and empirical studies of industrial conflict across time and nations*. Ph.D. Thesis, Industrial Relations Research Institute, University of Wisconsin, Madison, June.
- Shalev, M. (1979b). Strikers and the state: A comment. *British Journal of Political Science*, 8(4), 479–492.
- Shalev, M. (1983). The social democratic model and beyond: Two ‘generations’ of comparative research on the welfare state. *Comparative Social Research*, 6, 315–351.
- Shalev, M. (1990). Class conflict, corporatism and comparison: The Japanese enigma. In: S. N. Eisenstadt & E. Ben-Ari (Eds), *Japanese models of conflict resolution* (pp. 60–93). London: Kegan Paul International.
- Shalev, M. (Ed.) (1996). *The privatization of social policy? Occupational welfare and the welfare state in America, Scandinavia and Japan*. London: Macmillan.
- Snyder, D. (1975). Institutional setting and industrial conflict: Comparative analyses of France, Italy and the United States. *American Sociological Review*, 40(3), 259–278.
- Somers, R. H. (1971). Applications of an expanded survey research model to comparative institutional studies. In: I. Vallier (Ed.), *Comparative methods in sociology: Essays on trends and applications*. Berkeley, CA: University of California Press.
- Steenbergen, M. R., & Jones, B. S. (2002). Modeling multilevel data structures. *American Journal of Political Science*, 46(1), 218–237.
- Stimson, J. A. (1985). Regression in space and time: A statistical essay. *American Journal of Political Science*, 29, 915–947.
- Swank, D. (2002). *Global capital, political institutions, and policy change in developed welfare states*. Cambridge: Cambridge University Press.
- Swenson, P. (2002). *Capitalists against markets: The making of labor markets and welfare states in the United States and Sweden*. Oxford: Oxford University Press.
- Titmuss, R. M. (1974). *Social policy*. London: George Allen and Unwin (edited posthumously by Brian Abel-Smith and Kay Titmuss).
- Tufte, E. R. (1978). *Political control of the economy*. Princeton: Princeton University Press.
- Van Kersbergen, K. (1995). *Social capitalism: A study of Christian democracy and the welfare state*. London: Routledge.
- Verba, S. (1967). Some dilemmas in comparative research. *World Politics*, 20, 111–127.

- Visser, J. (1992). The strength of union movements in advanced capitalist democracies: Social and organizational variations. In: M. Regini (Ed.), *Labour movements towards the year 2000*. London: Sage.
- Weir, M., & Skocpol, T. (1985). State structures and the possibilities for 'Keynesian' responses to the Great Depression in Sweden, Britain, and the United States. In: P. B. Evans, D. Rueschemeyer & T. Skocpol (Eds), *Bringing the state back in*. Cambridge: Cambridge University Press.
- Western, B. (1996). *Causal heterogeneity in comparative research: A Bayesian hierarchical modeling approach*. Unpublished paper, Department of Sociology, Princeton University, December.
- Western, B. (1998). Causal heterogeneity in comparative research: A Bayesian hierarchical modelling approach. *American Journal of Political Science*, 42(4), 1233–1259.
- Wildeboer Schut, J. M., Vrooman, J. C., & deBeer, P. (2001). *On worlds of welfare: Institutions and their effects in eleven welfare states*. The Hague: Social and Cultural Planning Office.
- Wilensky, H. L. (1981). Leftism, Catholicism, and democratic corporatism: The role of political parties in recent welfare state development. In: P. Flora & A. J. Heidenheimer (Eds), *The development of welfare states in Europe and America* (pp. 345–382). New Brunswick, NJ: Transaction.
- Zorn, C. (2001). Estimating between- and within-cluster covariate effects, with an application to models of international disputes. *International Interactions*, 27(4), 433–445.