TOWARD IMPROVED USE OF REGRESSION IN MACRO-COMPARATIVE ANALYSIS

Lane Kenworthy

I agree with much of what Michael Shalev (2007) says in his paper, both about the limits of multiple regression and about how to improve quantitative analysis in macro-comparative research. With respect to the latter, Shalev suggests three avenues for advance: (1) improve regression through technical refinement; (2) combine regression with case studies (triangulation); (3) turn to alternative methods of quantitative analysis such as multivariate tables and graphs or factor analysis (substitution). I want to suggest some additional ways in which the use of regression in macro-comparative analysis could be improved. None involves technical refinement. Instead, most have to do with relatively basic aspects of quantitative analysis that seem, in my view, to be commonly ignored or overlooked.

LOOK AT THE DATA

Shalev's third suggested path for progress consists of using tables, graphs, and tree diagrams to examine causal hierarchy and complexity and to identify cases meriting more in-depth scrutiny. This should be viewed not as (or at least not solely as) a substitute for regression but rather as a critical component of regression analysis. All of us were (I hope) taught in our first

Capitalisms Compared
Comparative Social Research, Volume 24, 343–350

ISSN: 0195-6310/doi:10.1016/S0195-6310(06)24010-9

regression course that it is not enough to simply get the data, estimate some regression equations, and then draw conclusions. It also is necessary to get a feel for the data, in large part by examining descriptive statistics and looking at bivariate and/or multivariate patterns. Too many macro-comparativists, I suspect, either do not do this at all or do not do it sufficiently carefully.

In some instances what one finds by looking carefully at the data enhances or enriches what regression analysis tells us. Sometimes it calls into question the utility of using regression. Sometimes it suggests ways of altering the regression, for example by adding interaction effects, considering alternative functional forms of relationship, excluding certain cases, and so on. The tree diagram Shalev shows in discussing Bo Rothstein's (1990) analysis of determinants of unionization and the graph he uses in discussing Peter Hall and Robert Franzese's (1998) analysis of the impact of central bank independence and wage-setting coordination on unemployment are useful examples of what one can learn by spending a great deal of time looking at and thinking about the data. (That is not to suggest that either Rothstein or Hall and Franzese necessarily failed to do this. Sometimes we miss things, no matter how hard we look.)

It almost always is best to look at the data in graphical form. There are circumstances in which we can spot interesting patterns in tables. But it is much easier to do so when data are displayed graphically (Cleveland, 1993, 1994; Tufte, 2001; Wilkinson, 2001; Gelman, Pasarica, & Dodhia, 2002). Happily, these days the investment required to learn how to create both simple and relatively complex graphs is minor.

SHOW THE DATA

Quantitative analysis involves data reduction. But in my view, most recent quantitative macro-comparative work goes too far in this direction. The typical analysis includes 18 or so countries. In this type of research, unlike in analyses of thousands of individuals, the cases both are of substantial interest in and of themselves and can matter for our interpretation of regression findings.

The typical paper includes a few tables showing regression results and perhaps an appendix listing means and correlations among the variables. This is helpful information. But much more could be made available to readers. In particular, it is possible without taking up too much space to let readers see most of the raw data. In a cross-sectional macro-comparative paper the author can actually list all of the data used in the analyses – that is, the values for each country on each variable – in a table. For analyses that utilize longitudinal data, graphs displaying the time series for each country for key variables can be included. Bivariate or multivariate scatterplots can help readers (and authors) to see patterns in the data that warrant scrutiny.

GREATER TRANSPARENCY IN PRESENTING REGRESSION FINDINGS

In the typical textbook explication of multiple regression, the author shows the results of a bivariate regression (one independent variable), then introduces the notion of spuriousness and the concept of "controlling" with non-experimental data, and then proceeds to add one or more additional independent variables. The coefficient for the original independent variable changes, and the reader thereby learns about partial associations and omitted variable bias.

This analytical strategy is not only useful for pedagogical purposes. It is an appropriate way to proceed in "independent-variable-centered" analyses. In such analyses the research question concerns the effect of one (or sometimes two or three) independent variable on the outcome. The question is "what is the impact of X_1 on Y?" Sometimes, by contrast, the research is "dependent-variable-centered": the research question is "what causes Y?" In a dependent-variable-centered analysis it may be more appropriate to begin with a large number of (theory guided) independent variables and then gradually reduce the number according to criteria such as statistical significance or contribution to adjusted R^2.

Most analyses in macro-comparative research are independent-variable-centered. The question is something like "What is the effect of left government on social policy generosity?" or "What is the impact of wage-setting arrangements on unemployment?" Yet most analysts proceed by including as many controls as possible in their initial regression. Sometimes this is the only regression presented; in other instances some of the variables are then dropped and a second (and perhaps third and fourth) regression is shown.

A common circumstance is that we have fairly strong reason to suspect there will be an association between the hypothesized causal factor and the outcome, and the expected association is there at the bivariate level, but then it disappears in a multivariate analysis. Also common is that we have a not-terribly-compelling theory suggesting a link but no bivariate association, yet in the regression with 10 or so control variables the association appears.

Sometimes these multivariate findings are correct. But we should be suspicious. Those who have done enough multivariate regression analysis know well that it is sometimes (not always) possible to get the expected and/or hoped for finding to emerge if enough model specifications are tried.

As researchers and as consumers of others' research, we should want to know exactly how such a finding has emerged. That requires going step-by-step through the regressions, from bivariate patterns to the results of adding each of the various controls. Which particular control or set of controls makes the association change? Is that particular specification more theoretically compelling than others? How robust is the association to alternative specifications (not to mention measurement choices, groups of countries, and time periods)? Walk the reader through the analyses and findings. Allay suspicion by making it as transparent as possible what is going on in the data. Of course, space constraints typically permit showing only a limited number of the regressions. But the reader should nevertheless be informed of exactly what produced the result for the variable of interest in the preferred model specification.

WHICH VARIATION?

Macro-comparative analysts who use pooled cross-section time-series regression often fail to make clear what variation they aim to explain. There are three main options. One is variation in levels across countries. Here one can estimate cross-sectional effects averaged over multiple time periods (years, business cycles, decades). An example might be the impact of left government on welfare state generosity across 20 countries, averaged over the 1980s and 1990s. A second is variation over time within countries. Here regression can estimate an average over-time effect for a set of countries – for instance, the effect of left government on change in welfare state generosity in the 1980s and 1990s, averaged over 20 nations. A third is cross-country variation in change over time. We might, for example, be interested in the impact of left government on cross-country differences in change in welfare state generosity in the 1980s and 1990s.

Pooled regressions usually focus on one or the other of the first two of these, and most commonly on both. Following Larry Griffin, Walters, O'Connell, and Moor (1986) and Kittel (1999), Shalev rightly notes that a common problem with use of pooled regression in macro-comparative research is that researchers combine these two types of variation without (apparently) considering whether it is reasonable to expect that the causal process will be the same for both. Often that assumption is questionable.

Suppose cumulative left government is a major determinant of cross-country variation in welfare state generosity across 20 OECD countries as of 1980. But suppose it then has little or no effect on developments within these countries during the 1980s and 1990s. Perhaps over-time changes during these two decades are dominated by budget pressures and globalization. A pooled regression that does not distinguish between the determinants of cross-sectional variation vs. over-time variation will miss something very important in this type of situation.

Explaining cross-country variation in over-time changes is something different altogether. Suppose changes in budget pressures and globalization account for a significant portion of the longitudinal variation in welfare state generosity within each country in the 1980s and 1990s but that neither varies much across the countries. These two factors will not, then, help in explaining the differences between the countries in the direction and degree of over-time change. Those differences might instead be due to catch-up effects or to variation among the countries in public support for generous benefits or in the structure of the political system.

Fig. 1 illustrates these hypothetical differences in types of variation, using data on public social expenditures as a share of GDP. Setting aside the

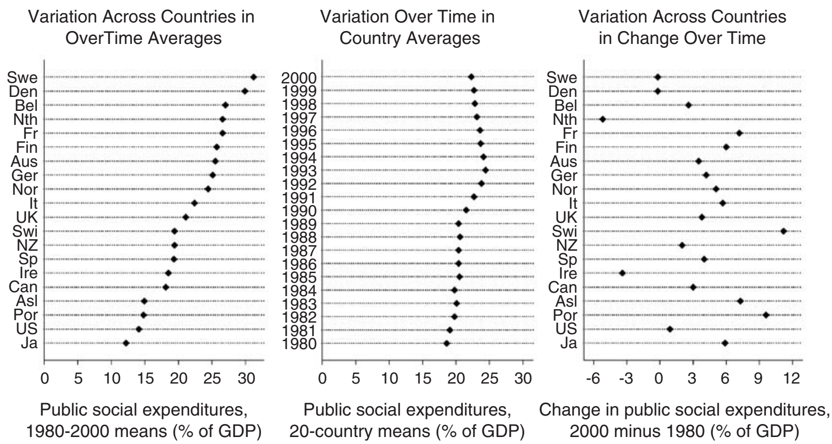

Fig. 1. Three Types of Variation in Public Social Expenditures. *Note:* Author's calculations from data in OECD (2004). The ordering of countries in the third chart follows that in the first chart, to highlight the contrast.

question of whether this is a useful measure of social policy generosity, the three charts show clearly that there are sizeable differences. This suggests the possibility of differing causal processes.

Macro-comparative researchers need to be clear about the type of variation to which their theory applies. And for empirical analysis the default assumption should be that causal patterns for cross-sectional and over-time variation differ. The utility of pooling should be demonstrated rather than presumed.

LONG-TERM VS. SHORT-TERM EFFECTS

Many pooled regression analyses use annual data. As best I can tell, most of the time that is because such data are available and because using them increases the number of observations, allowing for inclusion of more independent variables and enhancing statistical power. But many of the theories such analyses test imply medium-to-long-term effects. Sometimes analyses with annual data can pick up such effects, but that hinges on getting the lag structure correct. More often than not, using annual data to examine hypothesized medium-run or long-run associations will obscure rather than clarify.

But using longer time periods reduces the number of observations, heightening concern about omitted variable bias. What to do? There is no ideal solution. My preferred strategy is to examine all possible combinations of a "reasonable" number of independent variables (Kenworthy, 2004, 2007). For an N of 15 or so, that means perhaps three or four. This by no means eliminates worry about biased results due to improperly specified models. But inclusion of more independent variables is not inherently better in this regard (Lieberson, 1985; Achen, 2002). And in any event, having a better specification is not an improvement if the time period is wrong.

STATISTICAL SIGNIFICANCE?

Over the past decade much of the methodological debate in quantitative macro-comparative research has focused on how to properly estimate standard errors in pooled regressions. But in most instances such analyses include the full population of affluent countries in the time period considered. Where a sample is used, the sample is almost always dictated by data availability; there is no pretense that it is representative of the population.

Statisticians disagree about whether there is a rationale – based on the "superpopulation" notion – for considering statistical significance in this type of circumstance (Berk, 2004, offers a useful discussion). At the very least, however, analysts who believe standard errors are important to consider should offer an argument in favor of doing so, instead of simply doing so because it is conventional practice. Either way, many macro-comparative analyses would be substantially improved by paying more attention to the direction, size, and robustness of regression coefficients and less to statistical significance.

REGRESSION AS THE ANALYTICAL STARTING POINT

Because we often are dealing with the full population, macro-comparativists should treat analyses less as a means of drawing generalizable inferences and more as a means of understanding the cases (Ragin, 2001). In the prototypical quantitative macro-comparative article, the regressions are the starting and ending point of the analysis. I would like to see more papers in which regression is used to inform discussion of cases. What do the regression results tell us about why country A or regime-type B turned out as it did or changed in the way it did? Discussion of cases can then, of course, be used to question and/or further explore the regression results.

REFERENCES

Achen, C. H. (2002). Toward a new political methodology: Microfoundations and ART. *Annual Review of Political Science*, *5*, 423–450.

Berk, R. (2004). *Regression analysis: A constructive critique*. Thousand Oaks, CA: Sage.

Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ: Hobart Press.

Cleveland, W. S. (1994). *The elements of graphing data*. Summit, NJ: Hobart Press.

Gelman, A., Pasarica, C., & Dodhia, R. (2002). Let's practice what we preach: Turning tables into graphs. *The American Statistician*, *56*, 121–130.

Griffin, L. J., Walters, P. B., 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.), *Futures for the Welfare State* (pp. 101–138). Bloomington: Indiana 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*, 505–535.

Kenworthy, L. (2004). *Egalitarian capitalism*. New York: Russell Sage Foundation.

Kenworthy, L. (2007). *Jobs with equality*. New York: Russell Sage Foundation (forthcoming).

Kittel, B. (1999). Sense and sensitivity in pooled analysis of political data. *European Journal of Political Research*, *35*, 225–253.

Lieberson, S. (1985). *Making it count*. Berkeley: University of California Press.

OECD (Organization for Economic Cooperation and Development) (2004). *OECD social expenditure database: 1980–2001*. Paris: OECD.

Ragin, C. C. (2001). Case-oriented research. In: *International encyclopedia of the social and behavioral sciences* (pp. 1519–1525). Amsterdam: Elsevier.

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.

Shalev, M. (2007). Limits and alternatives to multiple regression in comparative research. *Comparative Social Research*, *24*, 259–308.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Wilkinson, L. (2001). Graphical methods: Presentation. In: *International encyclopedia of the social and behavioral sciences* (pp. 6368–6379). Amsterdam: Elsevier.