Step Away from the Pool

Lane Kenworthy
University of Arizona
lane.kenworthy@gmail.com

Quantitative macrocomparative analysis of the rich long-standing-democratic nations—"medium-N analysis"—is dominated by pooled time-series cross-section regression. I guesstimate that more than 90% of the medium-N papers I read in journals and as journal submissions use pooled regression.[1]

Quantitative data on many of the institutions, policies, and socioeconomic outcomes studied by comparativists first became available for more than a handful of countries in the 1970s. For a while quantitative macrocomparative research consisted mainly of cross-sectional analysis of single-point-in-time data or period averages. By the 1990s reasonably lengthy time series existed and analysts began examining over-time patterns. Pooling over time and across countries helped alleviate what had long been considered the achilles heel of cross-sectional comparative analysis: the small-N problem, which limits the number of control variables that can be included in a regression model. Pooled regression became the tool of choice. It has remained so for two decades.

In my view, that's unfortunate.

Better Apart

Aside from maximizing the number of observations, in many instances there is little reason to prefer pooling. Patterns of association across nations frequently differ from those over time. So why combine them? The default should instead be to examine them separately. This point has been made before—indeed, for quite a while now—so I won't dwell on it (Griffin et al. 1986; Kittel 1999, 2008; Kenworthy 2007, 2009; Shalev 2007).

But that leaves us with only 30 or so years, or only 20 or so countries. What about omitted-variable bias due to the small N? The degree of worry about this among comparativists is far out of proportion to its actual danger. Overspecified models are just as likely to mislead as under-specified ones (Lieberson 1985; Achen 2005), and the ability to throw in lots of controls tends to reduce the thought we put into our choice of models (Achen 2002).

Focus on Cross-Country Variation in Long-Run Change

Our aim should be to analyze changes rather than levels. Examining change gives us a better chance at identifying true causation.

Suppose we have data for 20 nations with annual observations from 1979 to 2007 (both years are business-cycle peaks). And suppose our interest is in the impact of a particular hypothesized cause. What type of analysis should we do?

The ideal is a differences-in-differences design, in which we compare changes over time across the countries. This is the closest approximation to an experiment. In an experiment we have two (or more) groups. We observe them at time 1. We administer the "treatment" to some observations. Then we compare the degree of change in both (or all, if there are more than two) groups. With observational data we don't administer the treatment, but in other respects the process is similar. Especially useful is that this design takes constant country-specific differences ("country fixed effects") out of play.

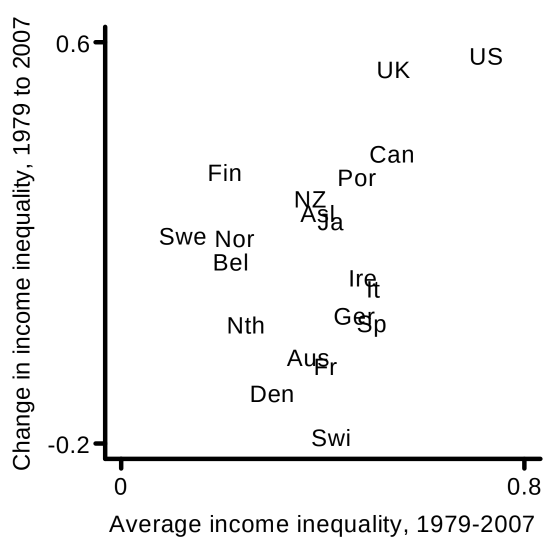
There remains plenty to worry about: selection into the treatment and control groups, reverse causality, country differences in other things that change, getting the lag right, variation in effects across subperiods. But this is a good design when appropriate.

When is it appropriate? When we have significant over-time change in the hypothesized cause, when there is variation across countries in that change, and when the change is mainly unidirectional rather than back-and-forth.

Doesn't this throw out a lot of interesting year-to-year variation? It does throw out that variation, but in many instances that variation has little or no bearing on our research question.

Here's an example. In recent years social scientists have grown increasingly interested in the effect of income inequality on socioeconomic outcomes such as health, crime, trust, and educational attainment (Burtless and Jencks 2003; Wilkinson and Pickett 2009). Inequality has risen significantly in some affluent nations. Figure 1 shows variation in income inequality change (vertical axis) and variation in income inequality levels (horizontal axis) for 20 countries. The range of values is the same on both axes.

Figure 1: Levels of Income Equality and Changes in Income Equality, 20 Countries, 1979–2007



Note: The income inequality measure combines the Gini coefficient for the bottom 99% of the income distribution and the top 1%'s share of income, with each rescaled to vary from 0 (smallest observation during the period) to 1 (largest). The same range of values, 0.8, is used on both axes. The change measure on the vertical axis is calculated as 2007 level minus 1979 level. Source: Jencks and Kenworthy, forthcoming.

Income inequality nicely fits the criteria for a differences-in-differences analysis. First, the amount of over-time change is substantial. Second, it varies a good bit across the countries. In fact, the cross-country variation in change between 1979 and 2007 is greater than in average levels over that period. Third, much of the change is unidirectional, as revealed by time plots for each of the countries (not shown here).

We could pool the data, using country-years or country-periods as observations, and do a pooled regression. But why? If a change in income inequality affects life expectancy, college completion, or trust, that causal process is likely to play out relatively slowly. Better, in my view, to use a simple differences-in-differences design, regressing change (first difference) in the outcome over the whole of the period on change in inequality.

This approach wasn't practical 20 or in some instances even ten years ago. The time series data didn't cover a long enough span of time. Now, for at least some interesting research questions, they do.

But Not Always

Next, consider an equally interesting and important question: What is the impact of tax levels on economic growth? Researchers have understandably—given the considerations noted earlier—tried to exploit the now fairly lengthy time series on tax revenues in OECD nations to get a handle on this question. Some analyses focus on the over-time variation within countries, some pool the over-time with the across, and some examine cross-country differences in over-time patterns.

A common tendency here, as in much macrocomparative research, has been to use yearly data. That makes little sense. The standard measure of taxation level is tax revenues as a share of GDP. This tends to move in a predictable and consistent way within business cycles, rising during growth phases and declining during recessions. This movement is of no relevance to the question of taxation's effect on economic growth. It's noise rather than signal. Our focus should be on the long run, not on fluctuations within business cycles.

How much variation is there in long-run change? Figure 2 does the same thing with tax revenues as a share of GDP that Figure 1 did with income inequality: On the horizontal axis is each country's average level during the period 1979–2007 and on the vertical axis is the amount of change between 1979 and 2007. The range of values is the same on both axes.

Here we see limited cross-country variation in change. More than two-thirds of the countries—those highlighted by the shaded horizontal band—experienced changes (increases) in tax revenues of less than five percentage points over these three decades. There was significant change in only four countries: Portugal, Spain, Italy, and the Netherlands. In a differences-in-differences analysis, this small set of countries is likely to drive the findings.

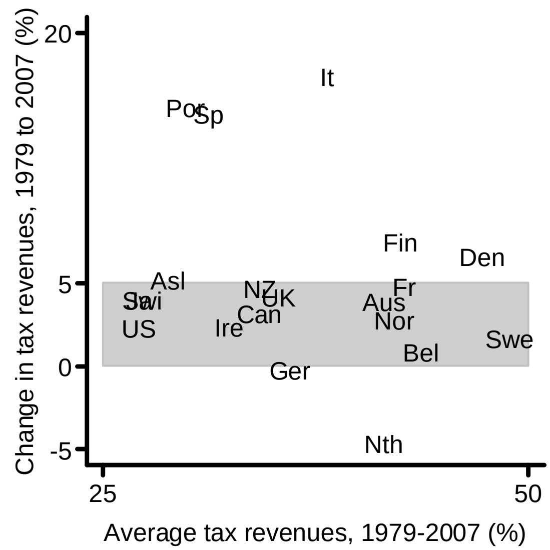
There is much more variation in average levels of taxation, shown on the horizontal axis. This looks to be a more promising source of information about the causal effect of tax levels on economic growth. It calls for a straightforward cross-sectional analysis of levels. We can regress average economic growth over 1979–2007 on average tax revenues over that period.

But doesn't that put us back in the small-N situation lamented by affluent-nation comparativists a generation ago? Yes, but the proper reason for lament was not the small N per se. It was the lack of sufficient data to know whether the analysis should instead focus on changes.

So what should we do? Everything we normally would do. Try some controls. Check for reverse causality. Consider influential cases. Look for interactions. Use case knowledge to ponder country fixed effects. Supplement the aggregate analysis with in-depth examination of a few countries.

Taxation isn't exceptional. There are lots of hypothesized causes in comparative politics and comparative political economy that have changed little over a fairly lengthy period of time: corporatism, cumulative left government, social policy generosity, welfare-state regime type, and health-care system, among others.

Figure 2: Levels of Taxation and Changes in Taxation, 20 Countries, 1979–2007



Note: Tax revenues as a share of GDP. The same range of values, 25 percentage points, is used on both axes. The change measure on the vertical axis is calculated as 2007 level minus 1979 level. Data source: OECD.

Summary

For the past two decades, a large amount of quantitative macrocomparative research on the world's rich nations has used pooled time-series cross-section regression with annual observations. I suspect this owes mainly to first-mover advantages and path dependence. In some instances, perhaps many, it's counterproductive. We almost always should separate the variation over time within countries from the variation in levels across countries.

Often we should focus on cross-country variation in long-run change. A differences-in-differences analysis is a good way to do this.

Sometimes there is little cross-country variation in changes, but quite a bit in levels. In such instances a cross-sectional analysis of levels is appropriate. It's rare to see this in published research these days. It probably strikes many scholars as primitive and archaic—a relic of a bygone era. That's a mistaken and unhelpful view.

Note

[1] I'm here referring only to papers that use solely country-level data. A growing number combine country and individual data and thus use a multi-level technique.

References

Achen, Christopher H. 2002. "Toward a New Political Methodology: Microfoundations and ART." *Annual Review of Political Science* 5, 423–450.

Achen, Christopher H. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22, 327–339.

Burtless, Gary and Christopher Jencks. 2003. "American Inequality and Its Consequences." In *Agenda for the Nation*. Henry Aaron, Pietro S. Nivola, and James M. Lindsay, eds. (Washington, DC: Brookings Institution), 61–108.

Griffin, Larry J., Pamela Barnhouse Walters, Phillip O'Connell, and Edward Moor. 1986. "Methodological Innovations in the Analysis of Welfare-State Development: Pooling Cross Sections and Time Series." In *Futures for the Welfare State*. Norman Furniss, ed. (Bloomington: Indiana University Press), 101–138.

Jencks, Christopher and Lane Kenworthy. Forthcoming. *Should We Worry About Inequality?* New Haven and New York: Yale University Press and Russell Sage Foundation.

Kenworthy, Lane. 2007. "Toward Improved Use of Regression in Macrocomparative Analysis." *Comparative Social Research* 24, 343–350.

Kenworthy, Lane. 2009. "The Effect of Public Opinion on Social Policy Generosity." *Socio-Economic Review* 7, 727–740.

Kittel, Bernhard. 1999. "Sense and Sensitivity in Pooled Analysis of Political Data." *European Journal of Political Research* 35, 225–253.

Kittel, Bernhard. 2008. "Statistical Narratives and the Properties of Macro-Level Variables: Labor Market Institutions and Employment Performance in Macrocomparative Research." In *Method and Substance in Macrocomparative Research*. Lane Kenworthy and Alexander Hicks, eds. (Basingstoke, UK: Palgrave Macmillan), 29–66.

Lieberson, Stanley. 1985. *Making It Count*. Berkeley: University of California Press.

Shalev, Michael. 2007. "Limits and Alternatives to Multiple Regression in Comparative Research." *Comparative Social Research* 24, 261–308.

Wilkinson, Richard and Kate Pickett. 2009. *The Spirit Level: Why Greater Equality Makes Societies Stronger*. New York: Bloomsbury Press.

Fast-Moving and Slowing-Moving Factors: Taking Apart Time-Series Cross-Sectional Analysis

Gary Goertz
University of Arizona
ggoertz@u.arizona.edu

In this essay I offer reflections on some basic methodological issues that all those in the field of advanced industrial-democratic societies face.[1] The scope of this research is typically limited to advanced industrial-democratic countries which number about 20–35 depending on the period. Time-series data is often available, depending on the country, for about 30–40 years (i.e., 1970s–2000s). This time-series cross-sectional situation forms one of two or three standard types, which also include many cross-sections and few time periods (e.g., panel survey data) or many cross-sections and many years (e.g., international conflict data which has many dyads, i.e., 20,000+, and many years, often 1816–2000). I also draw on the international conflict literature which faces many of the same problems and has discussed the same basic problem. The time-series cross-sectional data of comparative political economy faces many serious problems. There are various common solutions, including fixed effects for cross-sectional issues and lagged dependent variables and panel-corrected standard errors for temporal ones (see Beck 2001; Plümper, Troeger, and Manow 2005; and Kittel and Winner 2005 for good discussions in a comparative political economy context).

Inspired by Pierson (2004, chapter 3), we can think of the methodological issues that this field faces in terms of slow-moving (including stationary) versus fast-moving variables, as described in Table 1. In Table 1 I give some examples where the first variable is the independent variable and the second is the dependent.

Another way to think about fast versus slow moving is in terms of the location of the variance in the data. Fast-moving means more variation, while slow-moving means less. The extreme of slow-moving is constant, which is of course no variance at all. The extreme of fast-moving is that the data vary from one extreme to another every year. This provides a specific means for placing a given X–Y pair in one of the cells of Table 1.

My general proposition is that each of these cells has significant, but often different, methodological challenges. I focus mostly on the situation where one of the variables is slow moving, in part because I think the methodological challenges are more severe. As I discuss in more detail below, the standard methodology works best if the data (i.e., world) fits into the fast-fast cell. As such the most problematic cell for this methodology is the slow-slow one. The methodology for time-series cross-sectional is a "one size fits all." I think by default most scholars will analyze the data falling into these four cells in basically the same way. This is not to say that there is a consensus on how to analyze time-series cross-sectional data;