Statistics Corner

Time Series Analysis and Fractional Integration

By Professor Jan Box-Steffensmeier

Time series statistical techniques are used to evaluate social processes occurring through time. A basic definition of a time series is that it is a set of N time-ordered observations of a process where each observation is an interval-level measurement with equal measurements of time separating the observations. We are considering the case where T (time) clearly dominates N (number of observations). Below is an example of the type of data that would be analyzed. This is the time series for the partisan gender gap in the U.S. using quarterly data from 1977 to 2000. That is, the difference between the partisanship of men and women – aggregated by quarter with 95% confidence intervals.

The gender gap has ranged from nearly zero to just over 10% between 1977 and 2000. Women were about 5% more Republican than men in the late 1970s and were about 5-10% more Democratic than were men by the late 1980s and through 2000. The gap does not appear to grow steadily, but increases substantially in 1979 and keeps increasing at a slower rate. Since all surveys have margin of error, we include 95% confidence intervals (i.e., 1.96*margin of error).

While stark changes in the nature of the gender gap have stimulated individual level research explaining why men and women behave differently or have different preferences, the dynamics of the gender gap have gone largely unexplored. Time series techniques are uniquely suited and specifically designed to get at precisely these dynamics. It is clear that the gender gap is dynamic and importantly, that these dynamics exist outside of elections. The partisan gender gap has grown in a period of Republican strength, but appears to be maintained equally by similar movements toward the Democratic Party. Parallel and counter movements in men's and women's partisanship, different rates of decline and rebound, produce a gender gap that tells a more interesting story than that women identify more with Democrats than do men; *men and women react differently to the political environment*. But it is not clear what motivates these dynamics. What moves the partisan gender gap? What about the macropolitical environment has affected men and women differently? These are important questions both for improving our understanding of gender differences and for understanding the interplay of politics and gender as macropolitical events. Again, time series analysis is precisely the method to answer these types of substantive questions.

Examining the dynamics of the partisan gender gap gives us a unique perspective on the causes of gender differences and offers several advantages over cross sectional analysis. Time series analysis allows us to assess the role of the changing environment in determining gender differences in attitudes and behavior and to offer an explanation for this change. Specifically, by using time series data, we can test the hypotheses that changes are due to the effects of changes in the political environment; changes in the conditions or experiences of men and women; and changes in the economy. In so doing, we can better understand gender differences more generally. In contrast, cross sectional analysis is prone to several pitfalls. First, the persistence of the gender gap cannot be assessed cross sectionally. Second, explanations for the gender gap in cross sectional analysis focus on features that discriminate among individuals, which are inherently less political than macro level factors (Erikson, MacKuen, and Stimson 2002); political conditions are constant within a cross section and cannot help us understand variations in gender differences at any one point in time. Third, the factors that help us explain variation in individual behavior do not work well over time, in fact, they are usually constant. We, however, want to understand the impact of slow moving macro (social) processes, e.g., economic, social, and political processes. For example, the percent of women with a college education moves very slowly over time and is constant for all individuals if using a cross sectional approach. Finally, one cannot consider the evolution of the gender gap and associated counterfactuals over time. We want to be able to talk about the effect of a shock on the trajectory and sustainability of the gender gap. There is a role for both cross sectional and time series analysis, but they ask and answer different questions.

We assess the *persistence* of the gender gap because of its importance for thinking about shifts in aggregate partisanship. If gender differences are persistent, an increase in the gender gap due to shocks in the political, economic, and structural environment may last for years even if the shocks are just transient phenomena. In addition, theory predicts that gender differences will be persistent because they are a function of features of the political system that change very slowly. For example, it is difficult to move people off of welfare or change education levels, so gender differences will be persistent.

The level of persistence is determined by estimating the parameter, *d*, when characterizing a time series. When we relax the assumptions about *d* so that it can be any real number, we are using fractional integration. Fractional integration is an important innovation because of the increased precision of our estimates and characterizations of our series. If our series are not characterized appropriately, then subsequent analyses will not be accurate. Fractional integration also eliminated a major criticism of Box-Jenkins times series methods, specifically, the “art” of interpreting autocorrelation functions and partial autocorrelations functions to determine the characterization of the series was seen as too casual and was a controversial approach. Beran (1994) goes so far as to consider fractional integration a key unifying concept in time series analysis. Lanier et al. (1998) and Lebo (1998) argue that modeling data with fractional integration reduces spuriousness and improves parsimony. Importantly, fractional integration links macro and micro-level processes, which is theoretically important (e.g., Granger 1980; Box-Steffensmeier and Smith 1996). Finally, Lebo et al. (1998) show that most of the series of interest to political scientists are characterized by fractional dynamics, including presidential approval, consumer sentiment, macropartisanship, and ideology of the Supreme Court.

Fractional integration will be an integral topic in the upcoming time series course taught by John Freeman (University of Minnesota), Jon Pevehouse (University of Wisconsin and former OSU grad student) and myself. The course starts January 23rd and lasts for fourteen weeks. After a brief review of the calculus of finite differences and other estimation techniques, we study stationary ARMA models. In the next section of the course, we examine a number of important topics in time series analysis including "reduced form" methods (granger causality and vector autogression), unit root tests, near-integration, fractional integration, cointegration, and error correction models. Time series regression is also discussed (including a brief discussion of pooling cross sectional and time series data). We learn not only how to construct these models but also how to use them in policy analysis.

In contrast, Brian Gaines will teach an ITV course on the very large and rapidly growing statistical literature on analyzing data in which observations are a cross section of some units (e.g. individuals, countries, states, firms) over multiple time periods but where the cross sectional units clearly dominate. That is, the focus is on cases in which we have much larger Ns than Ts.

Here “panel data” is used in a broad sense to mean any data spanning multiple dimensions, usually (but not necessarily) time and space. In general terms, the advantage to having time and space variance is that one can avoid inferring that inter-personal differences across units are equivalent to inter-temporal differences within units. For instance, in a cross section if we find that age is a significant predictor of voting, we often infer that this result implies a forecast that all members of the relevant population will vote with increasing probability as they age. From a time series on just one unit, we might make the companion inference: that an observed life-cycle change in one unit implies across-age variation in a population that is heterogeneous in age in any given time period. Panel models are particularly prevalent in comparative politics.

Whether N (observation units) or T (time) dominate the data has a huge impact on the methods used. However, both time series and panel data methods take advantage of the inter-unit and inter-temporal variance. By using both dimensions of the data, one can typically adjudicate better among rival theories. The increasing availability of large data sets and powerful computers has made time series and panel models more and more prominent in the literature. For further information on both the time series and panel data courses and to obtain the syllabi, please see the ITV webpage at: http://psweb.sbs.ohio-state.edu/faculty/jbox/ITV/ITVHome.html