Social Science Research 45 (2014) 73–88

Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

Modeling time-series count data: The unique challenges facing political communication studies

Brian J. Fogarty [a], James E. Monogan III [b,*]

[a] Department of Political Science, University of Missouri—St. Louis, 800 Tower, St. Louis, MO 63121, United States
[b] Department of Political Science, University of Georgia, 413 Baldwin Hall, Athens, GA 30602, United States

ARTICLE INFO

Article history:
Received 25 March 2013
Revised 23 November 2013
Accepted 21 December 2013
Available online 9 January 2014

Keywords:
Political communication
Count models
Time series

ABSTRACT

This paper demonstrates the importance of proper model specification when analyzing time-series count data in political communication studies. It is common for scholars of media and politics to investigate counts of coverage of an issue as it evolves over time. Many scholars rightly consider the issues of time dependence and dynamic causality to be the most important when crafting a model. However, to ignore the count features of the outcome variable overlooks an important feature of the data. This is particularly the case when modeling data with a low number of counts. In this paper, we argue that the Poisson autoregressive model (Brandt and Williams, 2001) accurately meets the needs of many media studies. We replicate the analyses of Flemming et al. (1997), Peake and Eshbaugh-Soha (2008), and Ura (2009) and demonstrate that models missing some of the assumptions of the Poisson autoregressive model often yield invalid inferences. We also demonstrate that the effect of any of these models can be illustrated dynamically with estimates of uncertainty through a simulation procedure. The paper concludes with implications of these findings for the practical researcher.

1. Introduction

Political communication scholars commonly study an outcome that is a count variable, such as the number of stories printed on a subject in a given time frame (Flemming et al., 1997; Fogarty, 2005; Peake and Eshbaugh-Soha, 2008; Rhee, 1996; Sellers, 2000; Ura, 2009). Since the unit of analysis is a day, week, or month of coverage, these data often are time dependent as well. Many studies of such data focus primarily on specifying a reasonable time series model as if these outcomes were normally distributed, ignoring the count aspect of these data. Certainly, if we were limited to addressing only one of these issues, time dependence is the more important, as it speaks to non-independence of observations and more importantly the dynamic form of causality in time dependent data. (In other words, an input can have a consequence on the outcome many time periods into the future.) However, substantial developments have been made in the creation of Poisson models for time dependent data (Brandt et al., 2000; Brandt and Williams, 2001; Blundell et al., 2002; Schwartz et al., 1996), implying that models of media outcomes can handle both time ordering and the proper distribution for the outcome.[1]

In this paper, we illustrate how to apply time-series count models to media outcomes by replicating several important studies in the field of political communication. In so doing, we accomplish two things: We show how several results change under this new specification, and we illustrate how to interpret quantities in a count model for time series data. We begin this paper by describing the background on research about media event counts. Second, we describe various approaches to

* Corresponding author.
E-mail addresses: fogartyb@umsl.edu (B.J. Fogarty), monogan@uga.edu (J.E. Monogan III).

[1] For a more comprehensive review of time dependent count models, see Cameron and Trivedi (2013, Chapters 7 & 9) or Winkelmann (2008, Chapter 7).

0049-089X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ssresearch.2013.12.008

modeling time series of event counts, including the Poisson autoregressive model and its properties. Next, we conduct three replication analyses, describe how model-sensitive results are in each, and illustrate the best way to interpret these results. The three replications revisit Flemming et al.'s (1997) analysis of news stories about church and state issues, Peake and Eshbaugh-Soha's (2008) study of television coverage of energy policy, and Ura's (2009) study of newspaper coverage of homosexuality. We conclude by discussing the implications of our results for the practical researcher.

2. Background on models of event counts in political communication

As with most social science research, the sophistication of political communication analysis has been increasing over the past few decades. This includes an interest in modeling media count data over time. Instead of relying on descriptive measures alone, scholars are increasingly trying to systematically explain variations in coverage of political phenomena over days, months, and years (Flemming et al., 1997; Fogarty, 2005; Peake and Eshbaugh-Soha, 2008; Ura, 2009). At the most basic level, researchers are counting the number of articles, paragraphs, or television stories devoted to a subject in news outlets and then performing some type of multivariate analysis to assess what factors explain patterns in coverage.

We argue that researchers could use more effective techniques to analyze media count data over time. Time-series media count data are unique on several accords. Most obviously, models of count data rarely yield residuals that follow a normal distribution, and many models assume that population disturbances are normally distributed. It is generally more accurate to assume that count variables have a discrete distribution, such as the Poisson or negative binomial distribution, and model the mean of this distribution conditional on the predictors.[2] An even more important feature, though, is that these data are time-dependent. Recall that a Poisson process emerges when events emerge one at a time following an exponential distribution. One observation in a Poisson regression model offers the total number of events occurring in a set period. Further, independent increments of time are assumed, such the that number of events in one time period is independent of the event count in the other intervals (King, 1988, 839–841). This assumption is completely unfeasible for media count data, however, which can be cyclical or seasonal. One example is how the press covers national political institutions in the U.S. such as the Supreme Court, where coverage spikes during oral arguments and when decisions are handed down. More generally, a story printed today usually will have an effect on whether there is coverage tomorrow, the next day, or later. Exogenous events frequently dictate press attention, but reporters look to the past to explain the present and make predictions about the future. Therefore, we cannot model media count data as if stories are time independent.

There are two general ways most scholars treat media count data. One general strategy is for studies to ignore the dynamic element and pool the data, thus analyzing the observations as if they were cross-sectional (e.g., Barnhurst, 2003; Best, 2010; Bolsen, 2011; Farnsworth et al., 2010; Groeling and Kernell, 1998; Hayes and Guardino, 2010; Horvitz et al., 2008; Kahn, 1995; Morris and Clawson, 2005; Nisbet et al., 2003; Nisbet and Huge, 2006; Peake, 2007; Ramos et al., 2007; Ridout and Smith, 2008; Stack, 1990; Zaller and Chiu, 1996). In some of these studies, ignoring the time element is warranted as there is not enough variation in explanatory variables over the time span (e.g., Farnsworth et al., 2010; Ridout and Smith, 2008). At other times, it is simply not the goal of researchers to examine the time dynamics of news coverage (e.g., Bolsen, 2011; Ridout and Smith, 2008). Further, in many studies time is simply not the unit of analysis. This would be the case in any study treating events (such as Supreme Court decisions, prisoners' executions, or signing statements) as the unit of analysis with the next day's print coverage as the outcome. Time dependence therefore is not a concern if time does not define the unit of analysis. However, when time does define the unit of analysis, ignoring the time element diminishes our understanding of how the media cover certain political topics. As with most institutions, the news media are not static organizations, but instead reporting strategies evolve over time. Thus, explaining the media's coverage patterns must account for time.

Some specific examples include Horvitz et al. (2008), who examine press attention to the president's weekly radio address in the *New York Times*, *Boston Globe*, *Houston Chronicle*, and *Pittsburgh Post-Gazette* from 1982 to 2005. Presidential radio addresses are important to understand, as past research has shown that these addresses can affect presidential approval, which itself follows an autoregressive process (Baum and Kernell, 2001; Kriner, 2006). One of the dependent variables Horvitz et al. (2008) analyze in their valuable study is the number of paragraphs devoted to the weekly address in the newspapers. The authors ignore the time dependence of the coverage and simply pool coverage first by each president and then over the entire time span. Their results would be more convincing if they had modeled the likely autoregressive properties of coverage of presidents' weekly radio addresses; specifically, past addresses and coverage affecting current and future coverage. In this way, the model would allow factors shaping coverage to have spillover effects into future weeks and would account for non-independence of observations. In other important studies: Morris and Clawson (2005) examine coverage of Congress in the *New York Times* and the *CBS Evening News* in the 1990s, but ignore time dynamics that likely help explain patterns of coverage. Peake (2007) samples one day out of a week for five months in examining coverage of President Bush in 2006. Due to this collection of coverage over discontinuous time points, Peake also does not model time.

The second general strategy is for studies to model the time dynamics of their data but ignore the fact that the outcome variable is a media count (Barakso and Schaffner, 2006; Brown et al., 2001; Haider-Markel et al., 2006; Peake, 2007; Schreiber, 2010; Sellers, 2000). For example, Rhee (1996), recognizing the autoregressive issues with media count variables, uses the transfer function methodology (Box and Tiao, 1975; Box et al., 2008) to model the effect of polls on news coverage

[2] However, the distribution of counts does resemble a normal distribution with sufficiently large counts (Forbes et al., 2011).