Political Analysis (2004) 12:97–104
DOI: 10.1093/pan/mpg017

Time Series Intervals and Statistical Inference: The Effects of Temporal Aggregation on Event Data Analysis

Stephen M. Shellman
Department of Government, The College of William and Mary,
Williamsburg, VA 23187-8795
e-mail: smshel@wm.edu

While many areas of research in political science draw inferences from temporally aggregated data, rarely have researchers explored how temporal aggregation biases parameter estimates. With some notable exceptions (Freeman 1989, *Political Analysis* 1:61–98; Alt et al. 2001, *Political Analysis* 9:21–44; Thomas 2002, "Event Data Analysis and Threats from Temporal Aggregation") political science studies largely ignore how temporal aggregation affects our inferences. This article expands upon others' work on this issue by assessing the effect of temporal aggregation decisions on vector autoregressive (VAR) parameter estimates, significance levels, Granger causality tests, and impulse response functions. While the study is relevant to all fields in political science, the results directly apply to event data studies of conflict and cooperation. The findings imply that political scientists should be wary of the impact that temporal aggregation has on statistical inference.

1 Introduction

While a large body of literature in economics suggests that time series parameter estimates are sensitive to the unit in which the data are temporally aggregated, political science literatures are generally less concerned with the effects that aggregation has on coefficient estimates. That is, the temporal aggregation literature is largely ignored, as a whole, in political science journals. While many studies derive inferences from temporally aggregated data, such as those focusing on voter expectations, congressional budgets, arms transfers, business relations, international conflict and cooperation, and intranational conflict and cooperation, there is little emphasis on how temporal aggregation decisions affect the inferences we draw. There are of course some notable exceptions (Freeman 1989; Alt et al. 2001; Thomas 2002). The goal of this paper is to expand upon others' work on this issue and reexamine how aggregation affects our inferences. While the study should prove useful for all fields in political science, in particular, I wish to examine how temporal aggregation decisions affect inferences drawn from dynamic intranational conflict-cooperation time series models.[1]

Author's note: I would like to thank Will Moore and Sara Mitchell for their useful comments and suggestions. A more complete version of this paper is available on the *Political Analysis* Web site.

[1]Due to similarities with international event data analyses, the results of this study are directly applicable to time series analyses of both intranational and international conflict and cooperation.

Aggregating over time may intro bias.

Competing theoretical arguments are often tested head to head using temporally aggregated event data. Event data, which are "day-by-day coded accounts of who did what to whom as reported in the open press," offer the most detailed record of interactions between and among actors (Goldstein 1992, p. 369). To utilize event data in statistical models one must devise some method of aggregating different event types into a "single theoretically meaningful measure (in one or more dimensions)" of the relationships among actors (Goldstein 1992, p. 370). Most event data projects convert events into a measure of conflict or cooperation.[2] The conflict-cooperation variable measures the intensity of one actor's behavior directed towards another.[3] After combining the different event types into a conflict-cooperation measure, one must convert the events to a time series by temporally aggregating the data. To temporally aggregate data means to sum or average a variable across such regular time intervals as days, months, quarters, or years. The time interval is often chosen by the individual researcher (and should be theoretically motivated), but according to Alt et al. (2001), little attention has been paid to the bias that this choice can introduce.

While only a few political scientists examine the effects of temporal aggregation on inferences, many other scholars have explored the impacts of temporal aggregation in other literatures. In fact, when one surveys econometrics, economics, and statistics literatures, one generally concludes that aggregation matters.[4] That is, one can draw different inferences by choosing different units of temporal aggregation.

Marcellino (1999, p. 133) shows that "temporal aggregation usually alters most properties existing at the disaggregated frequency." As we aggregate data, we lose many of the statistical properties and dynamics of the raw series. Rossana and Seater (1995, p. 441) state that the most important effect of aggregation on time series data is the "virtual elimination of low-frequency variation. . . .Averaging . . . alters the time series properties of the data at *all* frequencies, systematically eliminating some characteristics of the underlying data while introducing others."

The general findings in economics and statistics suggest that temporal aggregation affects estimation and inference. However, most inquiry on this issue focuses on a particular model and problem within a specialized area of research. As such, I seek to determine if temporal aggregation affects time series empirical assessments of state-dissident conflict and cooperation. With regard to conflict-cooperation event data studies, some researchers aggregate the data by the day, others by the week, others by the month, and still others by the quarter. In this piece, I analyze Colombian event data in daily, monthly, and quarterly intervals and illustrate that aggregation decisions affect the inferences we draw from conflict-cooperation time series models.[5]

2 Hypotheses

I choose to posit a few general hypotheses that come from the repression and dissent literature that we can test using event data. I begin by hypothesizing that both states and

[2]Such projects include: Cooperation and Peace Data Bank (COPDAB), World Events Interaction Survey (WEIS), Integrated Data for Events Analysis (IDEA), Protocol for the Assessment of Nonviolent Direct Action (PANDA), and Intranational Political Interactions Project (IPI).

[3]See Azar (1982), Goldstein (1992), and Shellman (2003, 2004) for a more detailed discussion of the first issue related to conflict-cooperation scales.

[4]See Zellner and Montmarquette (1971), Wei (1982), Franzosi (1995), Granger and Siklos (1995), and Pierse and Snell (1995) for more on temporal aggregation and inferences.

[5]I also analyzed Afghan event data. These results are available in the longer version of this paper on the *Political Analysis* Web site.

States and dissidents drive each other's behavior.

Past behavior drives future behavior: policy inertia.

dissidents drive one another's behavior. We should see that states respond to dissident behavior and that dissidents respond to state behavior. In addition, we should see that states respond to their own past behavior and that dissidents respond to their own past behavior (i.e., the "policy inertia" hypothesis—see Goldstein and Freeman 1990, p. 23).

Next, I put forth two opposing directional hypotheses. First, the *backlash* or *escalatory* hypothesis states that when one actor increases its levels of cooperation (hostility), the other actor increases its levels of cooperation (hostility). That is, a positive linear relationship exists between the two actors' behaviors. Second, the opposing *deterrent* hypothesis posits that an increase in one actor's cooperative behavior will lead to the other actor decreasing its levels of cooperation (i.e., increasing conflict). That is, a negative linear relationship exists between the two actors' behaviors.

3 Research Design

I empirically test the above hypotheses in Colombia and Afghanistan but for the purposes of this article, I cover only the results for Colombia.[6] The Colombia data come from the Intranational Political Interactions (IPI) project.[7] The data are human coded using an ordinal scheme that contains ten general categories of cooperation and ten general categories of hostility (Leeds et al. 1995). To use statistical techniques, such as ordinary least squares (OLS) regression, the IPI event codes must be transformed into an interval-like measure of conflict-cooperation. To do so, I use the interval weights reported in Shellman (2004), which surveys expert domestic conflict scholars to produce an interval-like scale of conflict-cooperation for the IPI event data, where positive numbers indicate cooperation (>0 to +84.06) and negative numbers indicate hostility or conflict (<0 to −90.71). Once the interval-like conflict-cooperation data are obtained, the events must be converted to a time series by temporally aggregating the data.

Shellman used expert opinion to develop scale. Does he still have confidence in that technique?

The IPI project uses the day as the unit of observation. Therefore, the smallest temporal unit of aggregation that one could select with these data is the day. As such, I aggregate the data using the day. In addition, I aggregate over both the month and the quarter.

To test the hypotheses and assess the sensitivity of results to temporal aggregation, I specify an unrestricted two-actor vector autoregression (VAR) system of equations depicted here.

VAR System of Equations

$$S_t = \delta_{10} + \delta_{11}S_{t-1} + \cdots + \delta_{1n}S_{t-n} + \delta_{1n+1}D_{t-1} + \cdots + \delta_{1n+m}D_{t-m} + \varepsilon_{St} \tag{1.1}$$

$$D_t = \delta_{20} + \delta_{21}D_{t-1} + \cdots + \delta_{2n}D_{t-n} + \delta_{2n+1}S_{t-1} + \cdots + \delta_{2n+m}D_{t-m} + \varepsilon_{Dt}, \tag{1.2}$$

where S_t denotes state action taken at time t and D_t denotes dissident action taken at time t.[8] To specify the appropriate lag length, I estimated a series of VARs using a variety of lag lengths. I then compared the Schwartz Bayesian Criterion (SBC) from each model.[9] I chose the lag length that produced the smallest SBC across the various models.

OLS estimates are unbiased and asymptotically efficient because both Eqs. (1.1) and (1.2) have identical right-hand terms that consist of predetermined variables. However, there are other statistical problems encountered when estimating a VAR system of

[6]One can view Afghan results in the longer version of this article located on the *Political Analysis* Web site.
[7]To learn more about the IPI project, please see http://garnet.acns.fsu.edu/~whmoore/ipi/ipi.html.
[8]As such, all dissident groups are aggregated together to form one dissident actor's behavior, and all state actors are aggregated together to form one state actor's behavior.
[9]One may also choose to use the Akaike information criterion (AIC) to specify the appropriate lag length. I chose the SBC because it will always select the more parsimonious model (see Enders 1995, p. 88).

Potential problems.

equations. The first problem concerns regressing one nonstationary series on another nonstationary series. Doing so may produce a spurious relationship between the two variables and may result in falsely rejecting the null hypothesis. Therefore, one must check each time series that enters the VAR to see if it is stationary. I performed augmented Dickey-Fuller (ADF) tests on each temporal series in each VAR.[10]

The results of the ADF tests corroborate Pierse and Snell (1995), who find that unit roots are invariant to temporal aggregation. I find that each series is stationary across levels of aggregation with one exception, the quarterly Colombian dissident series. Since only one of the series is nonstationary, I run each VAR in levels (as opposed to differences). Sims (1980) recommends against differencing even if the variables contain a unit root. He argues that the goal of VAR is to uncover relationships among variables, not to produce meaningful parameter estimates. This leads us into our second problem.

A second problem concerns the effect that multicollinearity among the independent variables has on coefficient tests of statistical significance. Because the regressors are likely to be collinear, the t-tests on individual coefficients are not reliable (Enders 1995, p. 301). Therefore, I conduct block exogeneity tests (i.e., joint F-tests) to determine whether lags of one variable cause other variables in the system. To assess the direction of each relationship uncovered, I use vector moving average (VMA) methodology and plot the impulse response functions (IRFs). Just as an autoregression has a moving average representation, a VAR may be written as a VMA, in that the variables (S_t and D_t) are expressed in terms of their current and past values of the two shocks to the system (i.e., e_{St} and e_{Dt}).[11] Plotting the coefficients of the impulse response functions visually allows one to represent the behavior of the S_t and D_t series in response to various shocks. The simulations provide information on both the size and direction of the impact of each series, and thus provide information as to whether states' and dissidents' behavior is best characterized as policy inertia, reciprocity, or both.

4 Results

I present the results in three forms. To begin, the coefficient estimates are reported in Table 1 along with the joint F-statistics and Granger causality tests. The impulse response functions are presented in Fig. 1. As mentioned above, the t-statistics may not be reliable due to collinearity among the regressors. However, a few of the specified models contain only one lag. In this instance, we can directly interpret the coefficient of the single lag. In the case in which the model specifies multiple lags, we look to the block exogeneity tests to identify the reciprocal and policy inertia processes.

Differences in causation by aggregation method.

We begin by examining the F-tests in Table 1. The dissident daily series is characterized as policy inertia ($F = 38.15$), while the state series is affected by both policy inertia ($F = 3.29$) and responses to dissident behavior ($F = 10.86$). In contrast, at the monthly level, we find that state behavior "Granger causes" dissident behavior ($F = 3.19$), that dissident behavior cannot be characterized by policy inertia, and that state behavior is affected by neither policy inertia nor dissident behavior. In other words, we draw diametrically opposed conclusions using the daily data versus the monthly data. Finally, the quarterly series is not affected by past behavior of either actor.

The findings support Sims (1971) and Wei (1982), in that we do not infer the same "causal" relationships across temporal units of aggregation. Second, we see that as we

[10] A summary of the results for the ADF tests is reported in the version of this paper posted on the *Political Analysis* Web site.

[11] See Enders (1995, p. 305) for complete specification.

Table 1 VAR coefficients and block exogeneity tests: Colombia, 1983–1992

	Days		*Months*		*Quarters*	
Variable	D_t	S_t	D_t	S_t	D_t	S_t
S_{t-1}	.02	.05*	.41*	.08	−.009	.25
	(.03)	(.018)	(.165)	(.110)	(.316)	(.206)
S_{t-2}	−.05*	.00	−.02	.021		
	(.03)	(.018)	(.167)	(.110)		
D_{t-1}	.13*	.05*	.03	−.08	.27	.07
	(.018)	(.011)	(.109)	(.07)	(.212)	(.138)
D_{t-2}	.08*	.00	.19*	.11		
	(.018)	(.011)	(.105)	(.070)		
R^2	.03	.01	.04	.10	.06	.10
	Joint F-Tests		Joint F-Tests		Joint F-Tests	
$S_{t-1} = S_{t-2} = 0$	1.53	3.29*	3.19*	.27	–	–
$D_{t-1} = D_{t-2} = 0$	38.15*	10.86*	1.69	1.65	–	–
SBC	14.86		15.21		13.75	
df	3455		112		37	

S and D are state behavior and dissident behavior, respectively. The dependent variables are across the top and the independent variables appear in column one. All models are run with constants but those coefficients do not appear in the table. The lag lengths were specified using the Schwartz Bayesian criterion (SBC).
*Significant at the .05 level (one-tailed test).

move from lower levels (days) to higher levels (quarters) of temporal aggregation, the R^2 values generally increase (see Zellner and Montmarquette 1971). Finally, the results show that as we move from lower (days) to higher (quarters) levels of aggregation; relationships are masked and unexposed.[12] Before concluding, I examine the impulse response functions to uncover the direction of relationships and note the discrepancies across levels of aggregation.

The impulse response functions are used to conduct simulations in which one of the variables is shocked and the response of each of the other variables is traced over a given number of time periods. Figure 1 illustrates each IRF. Within each larger cell, the upper left and lower right graphs indicate one actor's response to itself, while the upper right and lower left graphs indicate one actor's response to the other actor. The y-axis represents an actor's behavior on a conflict-cooperation scale, where conflict is represented by negative values and cooperation is represented by positive values. The x-axis represents time. Note that the time period in each graph differs. In graph A, the x-axis represents days, in graph B, the x-axis represents months, and in graph C, the x-axis represents quarters. Each graph illustrates one variable's response (e.g., dissident behavior) to another variable (e.g., state behavior) over time, after introducing a one-standard-deviation increase in the independent variable (e.g., state behavior). In this study, a shock constitutes an increase in one actor's cooperation level. The dotted lines are confidence bounds. When a confidence bound contains zero, we accept the null hypothesis of no impact.

To begin, one should note that each series decays rather quickly due to the stationarity of the series. In each IRF, we see that the confidence bound contains zero after a brief number of time periods. A quick survey of the IRFs suggests that there are no sign changes

[12]See Goldstein (1991) and Franzosi (1995) for more on this finding.

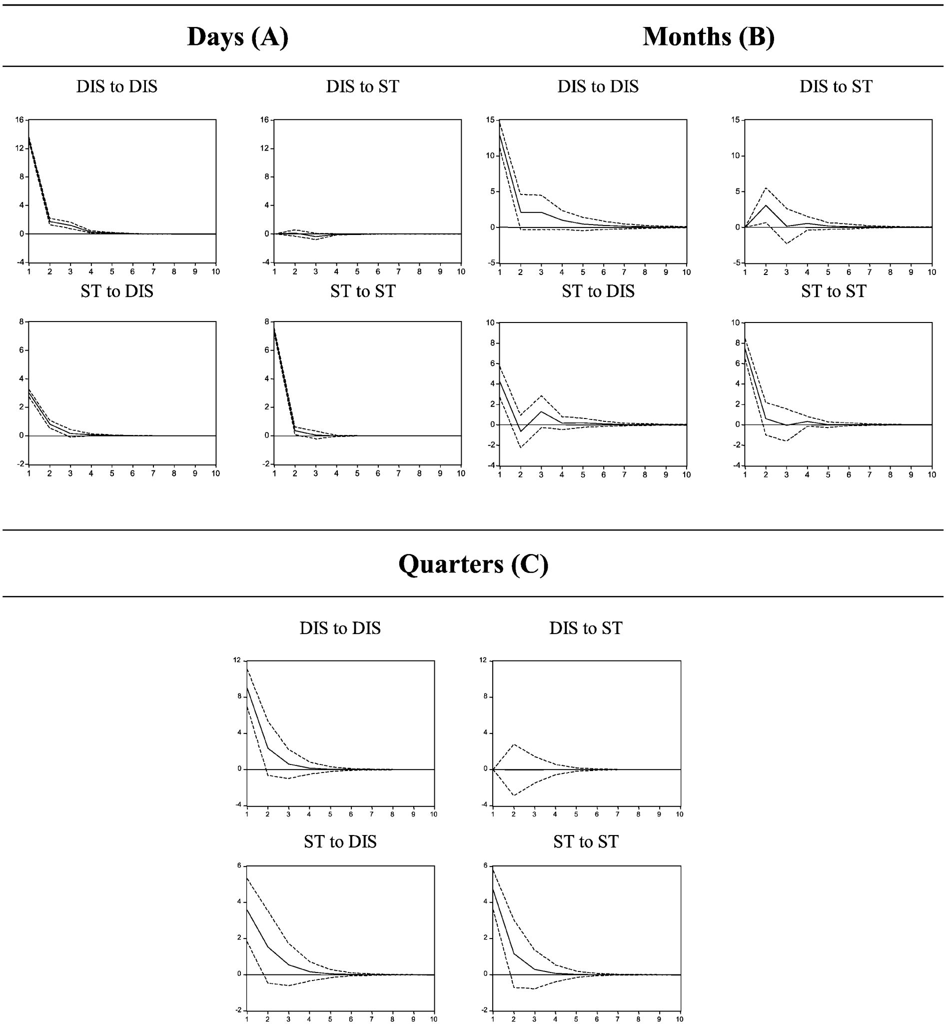

Fig. 1 Impulse response functions—Colombia. Each graph illustrates one actor's response (e.g., ST) to another (e.g., DIS) when a one-SD Innovation (± 2 SE) is introduced.

across aggregations. If there is a significant response, it is positive. The positive responses support the backlash or escalatory hypothesis. Because cooperation is positive, this may seem odd. However, if we were to introduce a negative shock, we would see a reversal of the IRF. The actor would respond in a conflictual or negative manner. Nevertheless, for our purposes, as a one-unit standard deviation of cooperation is introduced, the actor responds with cooperation.

Observe the IRF in quadrant A, in which the state is responding to the dissidents. In this instance we see that the state becomes more cooperative as dissidents become more cooperative. We observe this same effect in both the monthly (B) and quarterly (C) data. As far as the dissidents responding to the state, we pick this relationship up only in C (upper right IRF). There is no initial response, but by period 2, the dissidents are increasing cooperation levels. With regard to the impulse response functions, there does not seem to be much variation across the temporal units of aggregation.

5 Conclusion

The study demonstrates that temporal aggregation of intranational event data affects the inferences we draw from dynamic internal conflict-cooperation models and suggests that aggregation decisions may account for some of the mixed findings in the repression-dissent literature. So what should we do about it? To begin, we must put forth a theoretically driven argument about the temporal dynamics of state and dissident responses to one another. Rarely, if at all, do internal conflict scholars have a theory about the temporal frequency of conflict and cooperation between actors in a given case. If we are aggregating our data by a period that exceeds the appropriate interval, our empirical analyses "will be marred with temporal aggregation effects" (Zellner and Montmarquette 1971, p. 335). While aggregating data by the appropriate interval is plausible in many areas of study, the appropriate interval at which states respond to dissidents and dissidents respond to states is unclear. For example, budgetary processes that take place each year should be studied in annual units, but it is difficult, maybe near impossible, to derive an "appropriate" (i.e., natural) unit of time in which states and dissidents act and react to one another.

Some scholars have ignored the time period between events and sequentially aggregated event data using a *turn* metric (Dixon 1988; Moore 1998; Shellman 2003, 2004). A turn is a series of uninterrupted actions by one actor directed towards another. Turn data are employed to understand actors' sequential responses to one another, or what is often referred to as "action-reaction" processes. Depending on the question, we should continue to explore the utility of sequential models and perhaps abandon the use of discrete time metrics in some of our analyses.

However, in the absence of a theory that addresses the temporal dynamics of state and dissident reactions, we should conduct robustness checks across alternative levels of aggregation (Freeman 1989). The message from this study is that inferences are affected by aggregation decisions. Therefore, one should examine the sensitivity of results across various time intervals when assessing the tenability of knowledge claims founded in time series econometric studies. More weight should be given to results that hold across multiple levels of aggregation. Regardless, first and foremost, I would encourage scholars to consult theory when specifying the temporal dynamics of state-dissident interactions.

References

Alt, James, Gary King, and Curtis S. Signorino. 2001. "Aggregation among Binary, Count and Duration Models: Estimating the Same Quantities from Different Levels of Data." *Political Analysis* 9:21–44.

Azar, Edward E. 1982. *The Codebook of the Conflict and Peace Data Bank* (*COPDAB*). Unpublished. University of North Carolina at Chapel Hill.

Dixon, William J. 1988. "The Discrete Sequential Analysis of Dynamic International Behavior." *Quality and Quantity* 22:239–254.

Enders, Walter. 1995. *Applied Econometric Time Series.* New York: Wiley.

Franzosi, Roberto. 1995. *The Puzzle of Strikes: Class and State Strategies in Postwar Italy.* Cambridge: Cambridge University Press.

Freeman, John R. 1989. "Systematic Sampling, Temporal Aggregation, and the Study of Political Relationships." *Political Analysis* 1:61–98.

Goldstein, Joshua S. 1991. "Reciprocity in Superpower Relations: An Empirical Analysis." *International Studies Quarterly* 35:195–210.

Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of Conflict Resolution* 36:369–385.

Goldstein, Joshua S., and John R. Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics.* Chicago, IL: University of Chicago Press.

Granger, C. W. J., and Pierre L. Siklos. 1995. "Systematic Sampling, Temporal Aggregation, Seasonal Adjustment, and Cointegration: Theory and Evidence." *Journal of Econometrics* 66:357–369.

Leeds, Brett Ashley, David R. Davis, Will H. Moore, and Christopher McHorney. 1995. *The Intranational Political Interactions (IPI) Codebook.* (Available from http://garnet.acns.fsu.edu/~whmoore/ipi/codebook.html.)

Marcellino, Massimiliano. 1999. "Some Consequences of Temporal Aggregation in Empirical Analysis." *Journal of Business and Economic Statistics* 17:129–136.

Moore, Will H. 1998. "Repression and Dissent: Substitution, Context, and Timing." *American Journal of Political Science* 42:851–873.

Pierse, R. G., and A. J. Snell. 1995. "Temporal Aggregation and the Power of Tests for a Unit Root." *Journal of Econometrics* 65:333–345.

Rossana, Robert J., and John J. Seater. 1995. "Temporal Aggregation and Economic Time Series." *Journal of Business and Economic Statistics* 13:441–451.

Shellman, Stephen M. 2003. "Taking Turns: A Theory of Government-Dissident Interactions." Unpublished dissertation. Tallahassee: Florida State University.

Shellman, Stephen M. 2004. "Measuring the Intensity of Intranational Political Interactions Events Data: Two Interval-like Scales." *International Interactions* in press.

Sims, C. A. 1971. "Discrete Approximations to Continuous Time Distributed Lags in Econometrics." *Econometrica* 39:545–563.

Sims, C. A. 1980. "Macroeconomics and Reality." *Econometrica* 48:1–49.

Thomas, G. Dale. 2002. "Event Data Analysis and Threats from Temporal Aggregation." Presented at the Florida Political Science Association Meeting, Sarasota, FL, March 8.

Wei, William W. S. 1982. "The Effects of Systematic Sampling and Temporal Aggregation on Causality—A Cautionary Note." *Journal of the American Statistical Association* 77:316–319.

Zellner, Arnold, and Claude Montmarquette. 1971. "A Study of Some Aspects of Temporal Aggregation Problems in Econometric Analyses." *Review of Economics and Statistics* 53:335–342.