

A Solution to the Repeated Cross Sectional Design

Matthew Lebo
Stony Brook University
Matthew.lebo@sunysb.edu

Christopher Weber
Louisiana State University
crweber@lsu.edu

Abstract

Repeated (or “rolling”) cross-sectional (RCS) designs are distinguished from true panels and pooled cross sectional time series (PCSTS) designs by the fact that cross-sectional units – such as individual survey respondents – are not repeated at every time-point. Two serious problems pervade the use of RCS designs in social science. First, although RCS designs contain valuable information at both the aggregate and individual-level, available modeling methods – from pooled OLS to PCSTS to time series models – force researchers to choose a single level of analysis. Second, as with PCSTS, serial correlation is a serious problem. However, the two most common approaches to dealing with serial correlation in PCSTS data – differencing and using a lagged dependent variable – are not possible with RCS data because cases do not appear more than once in the data. Thus, the PCSTS toolkit does not provide a solution. We offer one here. Our method introduces the process of *Double Filtering* to cleanse the data of serial correlation and then uses multilevel modeling (MLM) to retrieve both aggregate-level and cross-sectional parameters simultaneously. The first of two filters estimates and fixes autocorrelation at the aggregate level using ARFIMA (auto-regressive fractionally integrated moving average) methods. A second filter, akin to mean centering in PCSTS designs, creates individual-level deviations from the aggregate-level model so that individual-level observations are also free of autocorrelation. We use Monte-Carlo experiments and the 2008 NAES to explore several modeling alternatives and demonstrate the supremacy of Double Filtering in a MLM-ARFIMA framework.

Introduction

There is an important distinction to be made between data-sets comprising the same observations over multiple time-points (true panels or pooled cross-sectional time series [PCSTS]) and those where the set of observations within each wave is not identical to the set of observations in other waves. The latter, what we may call “pseudo-panels,” have become increasingly common in political science and other social sciences.

Two types of pseudo-panel structures are distinguishable: the RCS design and the “unbalanced panel.” The unbalanced panel includes some units that appear in more than one time period but not all cases appear in every time period. For example, in a key piece on congressional elections, Canes-Wrone, Brady, and Cogan (2002) analyze House incumbents running for reelection over 21 congressional election cycles. The data are unbalanced in the sense that some members of Congress appear just once in the data, some many times over, but not all (or, in this case, any) cases appear in every cross-section. In political economy, Brown and Mobarak (2009) use an unbalanced panel where countries appear in their data over different portions of a 28-year period depending on data availability. Voeten (2008) studies decisions made by the European Court of Human Rights over multiple years but with turnover in the judges.

Honaker and King (2010) discuss the unbalanced panel as a missing data problem and provide a multiple imputation solution. This is an efficient and valuable method, especially useful for studies of international politics that see cases like developing countries enter and leave the data set at different times or where other gaps appear in available data. Yet, their solution cannot be applied to the situation where cases appear only once in the data – that is, when all but one observation of each case is missing in a dataset comprised of multiple time-points.

For this paper, we are interested in providing a solution for the other kind of pseudo-panel, the Rolling Cross Section (RCS). In an RCS design a unique set of cross-sectional units is measured at each time-point. Conceived of differently, a group of unique individual observations are divided into separate clusters and measured at different points in time.

The RCS data structure can be extremely useful by adding a dynamic component to the study of cross-sectional units. It gives researchers many of the benefits of traditional panel designs, less particular costs: problems of attrition and selection biases are not present because the same individuals are not tracked over time. Also, sample sizes in subsequent waves do not necessarily decrease. And, because the same unit is not tracked, there is not the problem of response bias, where answering a question at time $t-1$ influences how the question is answered at t . That is, familiarity with the study does not create a problem of correlated measurement error.

These designs have become more popular in recent years, in part, due to the many ways in which they can be created. The design is seen most frequently in survey research that employs similar survey items for new cross-sections over time (Gigengil and Dobrynska 2003). Examples include the National Annenberg Election Studies (NAES), the General Social Survey (GSS), and the cumulative American National Election Studies dataset (see, e.g: Stoker and Jennings 2008).¹ Many other RCS designs exist or can be easily compiled. The mass of data stored at ICPSR, for example, allows one to create RCS of CBS/NYT polls of monthly surveys going back to the 1970s measuring public opinion over time. The same can be done by collecting Gallup surveys, Princeton Survey Research Associates (e.g., Jerit, Barabas, and Bolsen 2006), Michigan's Surveys of Consumers (e.g., Clarke, Stewart, Ault and Elliott 2005), GIS, or the World Values Survey (e.g. Shayo 2009). For any of these RCS designs a wealth of observations can be studied for important relationships alongside

¹ The National Annenberg Election Study, for one, does this with repeated daily samples over the year prior to a presidential election. It draws a large random sample from the population and randomly splits the sample into replicates, which are contacted at a particular time during the campaign.

significant dynamics. Even these many RCS designs do not encompass the range of such data in use by political scientists.

In fact, during the 2004-2009 years alone, 68 articles using pseudo-panel data of some kind appeared in the *American Political Science Review* and the *American Journal of Political Science*. And, this number does not include designs that begin with individual-level survey data and aggregate them to create time series data (e.g., Box-Steffensmeier, De Boef, and Lin 2004).

Indeed, despite the breadth of RCS data, many seminal works in the discipline have begun by ignoring the valuable individual-level heterogeneity that exists within each time point, collapsing data sets into mean values for key variables, and examining the aggregate data using traditional time-series models at the daily, weekly, monthly, quarterly, or yearly level (among many, see: MacKuen, Erikson, and Stimson 1992; Romer 2006). For example, Johnston, Hagen, and Jamieson (2004) study dynamic campaign effects by aggregating responses over multiple days of the NAES, while examining the individual-level data in separate models (see also, Kensky, Hardy, and Jamieson 2010). Similar strategies have been used to study the gender gap in American politics (Box-Steffensmeier, DeBoef, and Lin 2004), consumer confidence (DeBoef and Kellstedt 2004), opinion change in response to political and social events (Green and Shapiro 1994), and Supreme Court decisions over time (Mishler and Sheehan 1993). The enormous body of literature on the dynamics of presidential approval and macropartisanship follows the aggregation strategy as a matter of course. Many of these studies can be considered foundation pieces in the study of American politics that improve our understanding of the movement of public opinion and electorates over time.

Studying such data in the aggregate has theoretical support. Kramer (1983), for one, argues that the actual state of the economy is an objective fact and that individual-level subjective

evaluations of it are either survey error or “partisanship, thinly disguised.”² Without disputing the value of aggregate studies, the aggregate- versus individual-level debate seems to be a false dichotomy given the possibility of multi-level models. Researchers should try more rigorous research designs rather than the entire avoidance of an important level of analysis. Basically, the solution of aggregating participants by day/week/month/quarter ignores within-time-point variation and can cut down datasets to a thousandth of their original size.

None of the pivotal aggregate studies mentioned has taken full advantage of the RCS framework where heterogeneity exists within as well as between time points.³ If important independent variables vary over time but are constant for a single time-point – e.g., the unemployment rate - there is a natural tendency to study data in the aggregate. But adding individual-level data in a multi-level model, as we will show, does not preclude the use of such variables. Researchers can complement aggregate studies and enhance our understanding of dynamic processes that use “long t ” time series by finding a role for individual-level data.

Without MLM, the individual-level approach can be taken too far, of course. Often, observations from various time-points are pooled (e.g., Romer 2006; Moy, Xenos, and Hess 2006; Stroud 2008; Kenski, Hardy, and Jamieson 2010). For example, Jerit, Barabas, and Bolsen (2006) pool dozens of public opinion surveys collected over a ten year time period to analyze the factors that influence political knowledge (see also Lau and Redlawsk 2008). This process of “naive pooling” treats the units as if they were collected in a single cross section (Stroud 2008). If observations within time points share unmeasured commonalities, this may lead to incorrect standard errors.

² Similar arguments can be found in MacKuen, Erikson, and Stimson (1989); DeBoef and Kellstedt (2004) and Box-Steffensmeier, DeBoef, and Lin (2004).

³ For example, Box-Steffensmeier et al. (2004) collect all available CBS/NYT individual-level survey data dating back to 1977 but aggregate them by quarter. In all, their time series data rely on the responses of over 250,000 unique individuals, yet the data they analyze consist of $n=87$ (p. 525).

Another method is to filter the time-component via fixed effects – effectively controlling for between-day effects. Doing this also has implications, in that it limits the researcher to explore static processes. It also assumes that parameter estimates pool around a common value. But are these approaches statistically sound and do they make the most of the data? As for the latter, it's evident that most published work using RCS has relied on statistical solutions that account for static *or* dynamic processes, not both. And, the statistical consequences of the various choices remain under-studied. Given the popularity of the RCS structure, this is disconcerting. Clearly, there is a need to understand the distinctive set of challenges RCS designs present and to explore the efficacy of several modeling choices – both old and new.

In this paper we first detail the unique aspects of RCS designs and discuss the problems with the most common modeling techniques that are used to deal with them. Following that, we outline our solution for dealing with RCS data – a two-stage model consisting of fractionally integrated time-series and multilevel modeling (MLM-ARFIMA). We then present Monte Carlo results that show the relative usefulness of several approaches and demonstrate the superiority of our method. Finally, we detail the results of this strategy in an applied example using the NAES.

Panels versus Pseudo-Panels: What's the Difference?

In a true PCSTS design, N units are observed over a fixed period of time, yielding an $N \times T$ size dataset. With such data, we are likely to encounter problems of autocorrelation in two directions. First, individual i at time t will be more correlated with individual j at time t than with individuals j at other time-points. Second, the values for each unit i are likely correlated with each other over multiple time-points. For example, in a Country by Year data set, the errors from a regression model will be prone correlation within years as well as for specific countries.

A key point is that, despite the fact that units are not repeated over time, neither of these two types of autocorrelation is any less likely in an RCS design. Autocorrelated errors again exist due to units being more highly correlated when observed at the same time. Neither does the problem of dynamic autocorrelation go away simply because units do not appear at multiple time points. Memory over time, traceable through aggregates, can exist and create autocorrelation between units more proximate to one another. That is, the error for i at time t is likely more correlated with that of j at time $t+1$ than with k at time $t+2$. The consistent finding of long memory in so many studies of aggregate RCS time series makes the possibility of serial correlation difficult to reject (see, e.g.: Box-Steffensmeier and Smith 1996; Lebo, et al. 2000).

Thus, the dynamic component poses particular problems in RCS data. So how are these problems to be dealt with? To begin, one should first realize the inadequacy of applying any of three common PCSTS approaches to the RCS case. With PCSTS (as well as traditional time series) including a lagged dependent variable (LDV) is a popular way to handle problems of non-stationarity (Keele and Kelly 2006). A second alternative, moving the LDV to the left-hand-side of an equation and differencing is also popular. By looking at the differences in observations between time points, a random walk series can be rendered stationary (Enders 2004).

Yet, if each individual observation occurs but once, these two approaches are simply impossible. That is, since individual i appears at time t but not at time $t-1$, $y_{i,t-1}$ does not appear in the data. Thus, using a lag as an independent variable is not a possible correction, nor is the use of a differenced dependent variable, $\Delta y_{i,t}$, created as $y_{i,t} - y_{i,t-1}$. A third solution, the use of panel corrected standard errors (Beck and Katz 1995), while a popular method for dealing with autocorrelation in PCSTS, is premised on observations repeating in every time point and does not solve the potential bias in coefficients. To this list of three, we could add Honaker and King's

(2010) imputation of data in “unbalanced panels” to resolve the problem of missing observations – a workable solution for a particular type of data, but not for the RCS design where each case will have data missing from every wave but one.

Modeling both dynamic and static processes together is a challenge with promise, so long as the results are reliable. This can be done in a multilevel framework and, within that structure, time series filtering techniques can correct for the problems presented by autocorrelation. We next turn to a discussion of MLM models and then to the specifics of our MLM-ARFIMA approach.

Modeling both Static and Dynamic Processes

Political scientists have increasingly relied on multilevel models to deal with hierarchical data structures in which “level-1” units are embedded or nested within “level-2” structures (Bartels 2009b; Gelman, Park, Shor, Bafumi and Cortina 2008). Often implemented in educational research (e.g., Bryk and Raudenbush 1992; Snijders and Bosker 1999) – the paradigmatic example being students nested within schools – as well as sociology (e.g., DiPrete and Grusky 1990), multilevel models provide a more holistic approach to analyzing hierarchical data, as they afford leverage to examine how contextual and individual level factors interdependently predict a dependent variable of interest (among others, Steenbergen and Jones 2002; Skrondal and Rabe-Hesketh 2004; Gelman and Hill 2007; Raudenbush and Bryk 2002). In addition to the substantive motivation to examine multiple levels of effects, there are also decisive statistical consequences that ensue when ignoring the hierarchical structure of a dataset. The problem is an error structure where observations are not independent. Insofar as observations are not independently sampled, but rather, drawn according to geographic areas or regions, for instance, the observed data will no longer be conditionally independent – that is, the errors will be spatially autocorrelated. As such,

the standard errors will be biased downwards and Type I error rates will increase (Skrondal and Rabe-Hesketh 2004).

The MLM, however, relies on the assumption that errors are both spatially and temporally independent. In an l -level model, the residuals are assumed to be conditionally independent at $l+1$. This assumption becomes tenuous with time in the model. Where errors are correlated over time, the standard errors for the model will be incorrect (Steenbergen and Jones 2004). Yet, the usefulness of the MLM model has led to several useful advances with data indexed over time.

Multilevel models have been used to analyze true panel data, where multiple observations are clustered at the country level (Beck and Katz 2007; Beck 2007; Shor, Bafumi, Keele and Park 2007). For example: $y_{i,t} = \alpha_i + \beta x_{i,t} + \varepsilon_{i,t}$, and, $\alpha_i = \gamma_1 + u_i$, where i may be a country level indicator for observations $1 \dots n$ observed repeatedly over time, t . In this case, the country is the level-2 unit, observed repeatedly over time (level-1). Beck and Katz (2007) note that if the assumption can be made that a dynamic process exists, a lagged DV can be included in the intercept equation. But, where the LDV is not measured, a different solution is needed.

Still, the problem can't be ignored since in RCS designs the standard errors may be incorrect due to clustering by interview date. It is identical to the problem of clusters in cross-sectional data – observations violate the assumption of being independently observed. Multilevel models are well-suited to deal with these data structures, as individual units can be viewed as embedded within the date the specific cross-section was collected (DiPrete and Grusky 1990).⁴

Our MLM-ARFIMA approach begins with the premise that individual-level data are nested within multiple, sequential time-points. Identical to clustering that is problematic in many cross-sectional datasets, the MLM can be thought of as a series of equations explaining the relationship

⁴DiPrete and Grusky (1990) advocate using a multilevel model to analyze RCS data but our method of double filtering is quite distinct from their approach.

between independent and dependent variables at increasing levels of aggregation (Gelman and Hill 2007; Skrondal and Rabe-Hesketh 2004). Here, the individual-level observations, i , are the level-1 units and are nested within level-2 units of time, t , be they days, months, quarters, etc.

It is important to note that using day-level variables (e.g., \bar{X}_t), and lagged day-level variables (e.g., \bar{X}_{t-1}), may not be enough to properly control for autocorrelation. This is where Box-Jenkins and fractional differencing techniques prove necessary (Box and Jenkins 1976; Hamilton 1994; Box-Steffensmeier and Smith 1996, 1998; Lebo, et al. 2000; Clarke and Lebo 2003). With Box-Jenkins techniques, short-term memory can be properly modeled with autoregressive and moving average parameters. And short-term processes can be modeled in an integrated series in an ARIMA framework. Where AR functions among cases (or similar *types* of cases) are heterogeneous, traditional differencing will be insufficient (Granger and Newbold 1974; Box-Steffensmeier and Smith 1996; Lebo, et al. 2000). By fractional differencing, the data generating process can be more accurately accounted for, white-noise is more easily produced, and the need for autoregressive and moving average parameters is reduced.⁵

We combine the logic of fractional differencing with multilevel modeling by first fitting an autoregressive fractionally integrated moving average (ARFIMA) model to the day-level RCS design.⁶ We then use a second filter for the individual data.

The important advances of our approach are several. First, we use the most reliable techniques available – (ARFIMA) models – to filter out autocorrelation present at level-2. Second, we take the deviations of i from level-2 values to fix problems of serial correlation at level-1. A third advance is that we are able to include level-2 variables that do not vary within time-points as

⁵However, additional parameters can be added after fractionally differencing the series.

⁶We discuss aggregation at the day-level to match the NAES example to follow. But these techniques are equally valid for data sets at other levels of aggregation such as monthly. Indeed, many of the findings of fractional integration in aggregate-level political variables are based on monthly or quarterly data (Box-Steffensmeier and Smith 1996; Lebo, Walker, and Clarke 2000). ARFIMA models can prove useful at any one of these levels of aggregation.

covariates. This allows us to explain together the dynamics occurring at level-2 and the important static effects at level-1.

To outline our model, we begin with level-2 and an equation familiar to time series researchers, the ARFIMA model:

$$(1 - L)^d \bar{Y}_t = \frac{(1 - \theta_q L^q)}{(1 - \phi_p L^p)} \varepsilon_t \quad (1)$$

where \bar{Y}_t represents the observed mean of all y_i within day t ; L is the lag operator such that $L^k Y_t = Y_{t-k}$; d is the fractional differencing parameter, the number of differences needed to render the series stationary; ϕ_p represents stationary autoregressive (AR) parameters of order p ; θ_q represents q moving average (MA) parameters; and ε_t is a stochastic error term for the level-2 disturbances.

By allowing values for d between 0 and 1 the series may be diagnosed as fractionally integrated (Box-Steffensmeier and Smith 1996). Yet, simpler models are available when d is an integer. Where $d=1$, the ARIMA model is produced with a differenced dependent variable:

$$\Delta \bar{Y}_t = \frac{(1 - \theta_q L^q)}{(1 - \phi_p L^p)} \varepsilon_t. \quad (2)$$

And, where the series is diagnosed as level stationary with $d=0$, a simple ARMA format suffices:

$$\bar{Y}_t = \frac{(1 - \theta_q L^q)}{(1 - \phi_p L^p)} \varepsilon_t. \quad (3)$$

The choice of which model to use – ARMA, ARIMA, or ARFIMA – will depend on the result of stationarity tests and the direct estimation of d where enough data are available for such tests (Enders 2004; Lebo, et al. 2000). Estimates of p and q can be obtained easily with integer values of d in any number of statistical software packages following the Box-Jenkins (1976) framework.⁷ The point of this step is to remove autocorrelation at level-2 so that an aggregate-level variable

⁷Estimating fractional values of d is somewhat more limited in terms of software but can easily be done in Stata, RATS, OX, and R.

can be explained by other factors aside from its own tendencies and past history. Estimates of (p,d,q) can be used to establish a noise model for \bar{Y}_t (Box and Jenkins 1976).

With these estimates one can apply the first of two filters. This fits within the Box-Jenkins framework of running a variable through its appropriate noise model to create a series of residuals that are devoid of autocorrelation:

$$\bar{Y}_t^* = (1 - L)^d \bar{Y}_t \times \frac{(1 - \phi_p L^p)}{(1 - \theta_q L^q)} \quad (4)$$

where \bar{Y}_t^* is just the residuals from \bar{Y}_t regressed on its noise model – a series that is both stationary in the long-run and free from autocorrelation due to short-run autoregressive and moving average processes (Box and Jenkins 1976; Box-Steffensmeier and Smith 1998). Put another way, \bar{Y}_t^* is \bar{Y}_t less its deterministic component, \bar{Y}_t' .⁸

For exogenous variables at level-2, a similar approach is then followed. When an exogenous variable varies only across time and not within a time period (e.g., the number of advertisements by a candidate) one should find the appropriate noise model for it and create Z_t^* , the deviation from Z_t not due to the past history of Z .⁹ Where exogenous variables vary within each day, means should be calculated and noise models created for each \bar{X}_t . This will allow the construction of \bar{X}_t^* which, along with \bar{Y}_t^* and Z_t^* means that level-2 is cleansed of autocorrelation.

Next, a second filter subtracts the daily deterministic component from the level-1 dependent variable:

$$y_{it}^{**} = y_{it} - (\bar{Y}_t - \bar{Y}_t^*) \quad (5)$$

⁸ This follows from \bar{Y}_t being a function of two components: $\bar{Y}_t = \bar{Y}_t^* + \bar{Y}_t'$. Subtracting out the deterministic portion of \bar{Y}_t , \bar{Y}_t' , from \bar{Y}_t removes the influence of the past history of \bar{Y} .

⁹ To distinguish the two types of exogenous variables we use Z for those that vary only over time and X for those that vary within a time-point as well.

And, where level-1 variation in the covariates exists, one should employ cluster mean-centering, a common practice in PCSTS and MLM (Baltagi 2005; Bafumi and Gelman 2007). Simply centering level-1 data around the within-day means removes the problematic day-level variation:

$$x_{it}^{**} = x_{it} - \bar{X}_t \quad (6)$$

The logic is the same as examined by Bafumi and Gelman (2007). By accounting for level-1 and level-2 effects, correct parameter estimates can be retrieved.¹⁰ The additional problem in the RCS design is that by failing to account for the autocorrelation brought about by time series data, one may reach erroneous conclusions.

The MLM now puts these double-filtered data to work.¹¹ The level-2 equation can include covariates that vary either strictly between days (Z) or both between and within days (X):

$$\bar{Y}_t^* = \alpha_2 + \beta_2 \bar{X}_t^* + \gamma Z_t^* + u_{2t}. \quad (7)$$

The level-1 equation provides the model of within variation:

$$y_{it}^{**} = \alpha_1 + \beta_1 x_{it}^{**} + u_{1it}. \quad (8)$$

where u_{2it} and u_{1it} are the respective errors for the level-2 and level-1 units.

While the process of double filtering comes before estimation of the MLM, equations (7) and (8) can be estimated together, combining both the within and between day effects:

$$y_{it}^{**} = \alpha_1 + \beta_1 x_{it}^{**} + u_{1it} + \beta_2 \bar{X}_t^* + \gamma Z_t^* + u_{2t}. \quad (9)$$

¹⁰ To obtain “within day” deviations, we remove the random and non-random variation in \bar{X}_t , where $\bar{X}_t = \bar{X}_t^* + \bar{X}_t'$. Thus, $x_{it}^{**} = x_{it} - (\bar{X}_t - \bar{X}_t^*) - \bar{X}_t^* = x_{it} - \bar{X}_t$.

¹¹ This approach is distinct from the “differences-in-differences” (DID) model frequently used to analyze RCS data structures, where cross-sections are included before and after a policy intervention (Wooldridge 2001; Heckman and Payner 1989; see Athey and Imbens 2006 for a thorough review of this literature). The DID may be used to test an intervention in populations affected and unaffected by the intervention. As such, the model simultaneously controls for aggregate effects, population differences, and the effect of an intervention by comparing two populations. As empirically flexible as the DID model is, it is important to underscore how it is different from the approach we advocate: first, it requires a discrete intervention; second, time is a discrete variable, often modeled by a dummy variable denoting whether the unit was observed before or after the intervention. Thus, the DID method does not afford immediate leverage to explore more dynamic processes, such as whether the clustered data follows a particular autoregressive pattern.

In (9) y_{it}^{**} is the double filtered values for y_{it} , which is a function of level-1 x s, aggregate level white-noise X s, covariates at level-2, and error components that vary within and between days.

One additional option that MLM models are well suited for is the estimation of time-varying parameters. As we do below in our NAES example, one can specify coefficients that will vary across time for certain independent variables, w_{it}^{**} . If a level-1 relationship might change across waves, a time varying coefficient, δ_t , can be specified. Thus, Equation (8) can be expanded to:

$$y_{it}^{**} = \alpha_{1t} + \beta_1 x_{it}^{**} + \delta_t w_{it}^{**} + u_1. \quad (10)$$

In sum, the steps can be outlined as follows: first, create means for each day of the level-1 variables of interest, \bar{Y}_t and \bar{X}_t . Second, find the appropriate noise models for the series of means, \bar{Y}_t and \bar{X}_t , as well as for the level-2 variables, Z_t , that do not vary within days and then, third, filter each series through its noise model. This will create \bar{Y}_t^* , \bar{X}_t^* , and Z_t^* , level-2 variables free of autocorrelation. Fourth, remove the day-level deterministic component from the individual-level data. Fifth, estimate the MLM in two levels using the doubly filtered data. To demonstrate the efficacy of our approach, we use Monte Carlo analyses to compare our approach with several alternatives for RCS data.

Simulations

The statistical consequences of various approaches to using RCS data are unclear. We expect that if the dynamic component in the data is ignored, parameter estimates will be adversely affected. And, we expect that parameter and standard error estimates will suffer when there is greater time dependence at level-2. Further, even if a lag at level-2 is modeled, bias will still be present to the extent that the lag does not completely account for autocorrelated errors that may exist within cross-sections (Keele and Kelly 2006). Moreover, if individual observations at time t

are more correlated with one another than with observations at $t+s$, there is a problem of clustering in the data; the errors will not be independent and the standard errors will be incorrect.

To further illuminate both the statistical problems and solutions, we simulate data meant to mimic the properties of RCS data. We generated 11,000 data sets (1,000 datasets per level of d) with each consisting of 275 waves with a sample size of 100 per wave.¹² Aggregate values of the independent variable, \bar{X}_t^* , were created along with x_{it}^{**} values for within-day variation.¹³ Level-1 observations, y_{it} , were generated as a function of within day effects, x_{it}^{**} (specified to have a slope coefficient of 0.5), between day effects, \bar{X}_t^* (specified to have a slope coefficient of 0.3), and random error.¹⁴ Next, series for \bar{Y}_t and \bar{X}_t were calculated so that there were 1,000 data sets for each value of fractional integration between 0 and 1 in increments of 0.1.¹⁵

We tested the statistical properties of eight estimation strategies for each data set. We start by presenting “naïve models” – models where a researcher would fail to separate out the between and within day effects. We do this using (1) OLS – labeled here OLS-Naïve – as well as (2) a multilevel model – MLM-Naïve – where intercepts vary across days. Next, we report the consequences of six additional estimation strategies that could feasibly be used with RCS data: (3) OLS pooling all data, but separating between and within day effects (OLS), (4) OLS, specifying

¹² Our datasets actually begin with 300 cross-sections, but we allow the first 25 to serve as a “burn in” for our models, since establishing the memory in the first few sets of observations is problematic.

¹³ Day-level means of x are drawn from a standard normal distribution. We then duplicate these observations 100 times to generate a dataset of size 27,500. These observations serve as the day-level random noise (\bar{X}_t^*). Next, we take a random draw from a standard normal distribution of size 27,500. These observations serve as with within day independent variable, x_{it}^{**} .

¹⁴ We added error to the model in two places, specifying a “within day” error distribution that is normally distributed with mean 0 and variance 4 and a “between day” error distribution of mean 0 and variance of 4. This ensures a large, but reasonable, relationship between x and y . It is important to note that we varied the ratio of between-day to total variation, or the “intra-class correlation.” Even with a small intra-class correlation ($\rho = 0.01$), the multilevel model we advocate outperforms the alternatives.

¹⁵ To do this, we first calculated the day level means for y and subtract y_{it} from these values. This gives us the deviations from the day level mean. Then, we fractionally integrated the day level means and added back in the deviations, which gives the value of y_{it} that one would observe. We followed the same process for x_{it} . The only thing that we vary in the simulations presented is d , the degree of fractional integration.

between and within day effects and including a day-level lagged dependent variable (OLS-LDV), and (5) OLS, accounting for non-stationarity by fractionally differencing the day level means (OLS-ARFIMA). We also estimated three additional types of multilevel models to account for unobserved heterogeneity across days – specifically, (6) MLM, separating between and within day effects and allowing intercepts to vary across time (MLM), (7) MLM, again separating between and within day effects, allowing intercepts to vary across time and including a day-level lag (MLM-LDV), and (8) MLM, fractionally differencing the aggregate series and allowing intercepts to vary across time (MLM-ARFIMA). All simulations and statistical tests were carried out in *R*.¹⁶

Simulation Results

Naïve Models. To demonstrate the consequences of ignoring day level effects and simply regressing y on x we generate two naïve models, one using OLS and a second using a MLM where intercepts vary across units. Figure 1 demonstrates the empirical consequences of these strategies. The upper panel displays the estimated slope for OLS and the lower panel for MLM estimates. The blue line is the true “within-day” slope and the red line is the “between-day” slope. The dots represent the estimates from the simulated datasets, with the solid black line representing the average of estimates at each level of d . For OLS, the estimates fall between the true slopes with low levels of d . But as d increases, so does the spread of the estimates and the average size of the estimated coefficient is biased towards zero. In other words, as d becomes less stationary the estimated slopes are more biased and less efficient.

– Figure 1 about here –

The bottom panel of Figure 1 illustrates the retrieved slope from the MLM naïve approach. The method properly retrieves the within-day effect but the empirical limitations of the strategy

¹⁶ The MLM models were estimated using the `lmer()` function in the “lme4” package (Bates and Maechler 2010).

are twofold: (1) it does not allow one to effectively model day-level processes, since within and between day effects are inseparable (see also, Bafumi and Gelman 2007; Skrongdal and Rabe-Hesketh 2004; Bartels 2009a), and: (2) it gives incorrect standard errors. As d increases, the standard errors will be biased downwards, leading to incorrect inferences.

Moving beyond the naïve models, we need to confront the problems of modeling both within- and between-day effects together as well as address the likelihood of non-stationarity at level-2. The latter problem is one that has been especially ignored by social scientists.¹⁷ Extending research that advocates separating cluster and within cluster effects (Bafumi and Gelman 2007; Skrongdal and Rabe-Hesketh 2004; Bartels 2009a), we explore the empirical consequences that ensue when these clusters are non-independent – i.e., they are not level-stationary. We examine six additional approaches – OLS, OLS-LDV, OLS-ARFIMA, MLM, MLM-LDV, MLM-ARFIMA.¹⁸

What should we expect from each approach? By pooling all observations and running an OLS regression, (solution 3: OLS), we should retrieve incorrect parameter estimates for the “between day” effect of x (\bar{X}_t). The fourth approach – simply specifying an OLS model with a lagged day-level dependent variable, \bar{Y}_{t-1} (OLS-LDV), should only produce unbiased and efficient estimates if the lag accounts for day-level autocorrelation (Achen 2000). Since this will not occur in the presence of fractional integration, the parameter estimates will be biased. The third approach, specifying an ARFIMA model for day-level x (\bar{X}_t) and day-level (\bar{Y}) and employing the method described in Equation 8 with OLS will result in incorrect standard errors, since OLS cannot effectively account for unobserved day-level variation.

¹⁷ For instance, researchers studying campaign effects have merged opinion data with spending (Kenny and McBurnett 1992) and advertising data (Freedman, Franz and Goldstein 2004) to examine the consequences of aggregate variables on voter decision making and behavior.

¹⁸ We estimated d in our models using Hurst R/S statistic (Hurst 1951). We estimated d in other ways. Our analysis revealed that the Robinson's estimator and the Geweke-Porter-Hudak (GPH) were slightly inferior to R/S, so we opted to use this estimate. The value of d = the Hurst R/S coefficient minus 0.5.

The MLM approaches should be an improvement over OLS by accounting for the clustering in the data. However, an assumption of the MLM is that level-2 errors will be independently distributed, which is violated insofar as level-2 units are correlated. Thus, the sixth solution, MLM, will produce biased and inefficient estimates as d increases. Similarly, MLM-LDV – the multilevel model with a level-2 lagged dependent variable – will produce biased estimates with standard errors that are increasingly incorrect as d increases. We expect that the MLM-ARFIMA model that merges the fractional differencing approach with the multi-level model will account for all these problems and prove to be the most reliable approach.

Figures 2 and 3 display our estimates of bias and inefficiency for the various OLS and MLM models, respectively. Bias was calculated by dividing each estimated parameter estimate by the true parameter estimate, calculating the average for each level of d and then multiplying by 100. Thus, values of 100 indicate a lack of bias. The degree of variation around the average estimate was calculated using the root mean squared error $RMSE = \sqrt{\frac{\sum(\theta - \bar{\theta})^2}{n}}$, where n is the number of replicated datasets for each value of d (i.e., 1,000). A small RMSE is preferred over a large RMSE, as it indicates less variation around the average estimated value. We display the results for four sets of estimates: bias and RMSE for each of the between-day effect (β for \bar{X}_t for the OLS, OLS-LDV, MLM, and MLM-LDV models and β for \bar{X}_t^* for the OLS-ARFIMA and MLM-ARFIMA models) and the within-day effect (β for x_{it} for the OLS, OLS-LDV, MLM, and MLM-LDV models and β for x_{it}^{**} for the OLS-ARFIMA and MLM-ARFIMA models).

– Figures 2 and 3 about here –

Figure 2 demonstrates that OLS and OLS-ARFIMA perform reasonably well in terms of retrieving the correct slope for the between-day effect of x on y . OLS-ARFIMA is the most accurate, which is to be expected since it effectively controls for non-stationary day level effects. OLS-LDV,

however, is increasingly ineffective as the estimated slopes are biased downwards as d increases. The upper-right quadrant similarly shows that OLS-ARFIMA has no problems of inefficiency while OLS and OLS-LDV grow more inefficient as d increases. All three methods perform well in terms of retrieving correct within-day effects, evident in the fact that the lines can barely be discerned in the bottom left quadrant of Figure 2. So long as one subtracts the day-level means from the observed data, correct within-day parameter estimates can be retrieved. But, the efficiency of estimates is compromised in the case of the OLS models. This can again be seen in Figure 4 which presents the distribution of standard errors for the three OLS approaches at three values of d .

Comparing Figure 2 to Figure 3, there are negligible differences. MLM-ARFIMA estimates of β for \bar{X}_t^* are unbiased and efficient, outperforming MLM and MLM-LDV. The distribution of MLM standard errors is shown in Figure 5 with the MLM-ARFIMA approach clearly standing out as best. All the results demonstrate the importance of accounting for non-stationarity.

– Figures 4 and 5 about here –

As one final check on the models, we investigate the observed differences in the standard errors across these models and for various degrees of non-stationarity. To this end, we present “optimism” in Table 1, which contrasts the estimated standard errors to sampling variation (Beck and Katz 1995; Shore et al 2007). In line with Beck and Katz (1995), we calculate optimism as follows: $Optimism = 100 \times \sqrt{\frac{\sum_{l=1}^{1000} (\theta_l - \bar{\theta})^2}{\sum_{l=1}^{1000} SE \theta_l}}$. Values greater than 100 indicate that true sampling variation is greater than estimated variation and standard errors are too small; values less than 100 indicate that standard errors are too large, since true sampling variation is smaller than estimated variation (Beck and Katz 1995). Thus, values in excess of 100 increase the probability of Type 1 error, rejecting a null hypothesis when it is in fact true.

As Table 1 illustrates, the standard errors are much too small for *all* methods except the MLM-ARFIMA model.¹⁹ At all levels of d , the standard errors are severely “over-confident” in any of the OLS models. That is, true sampling variability is much larger than estimated variability and the increase will lead to t-statistics that are inappropriately large. For the OLS and OLS-LDV estimates this effect is exacerbated as d increases. This unacceptable increase is also evident for the MLM and MLM-LDV models. As the day-level means are increasingly a function of past values, the standard errors are underestimated.

The OLS-ARFIMA models do well at various values of d but have optimism scores that are consistently high due to the model’s inattention to level-2 heterogeneity. Thus, Table 1 clearly illustrates that the only reliable approach with respect to optimism is the MLM-ARFIMA model. Only after accounting for fractional integration in the data, as well as unobserved level-2 heterogeneity, can one retrieve standard errors that more closely mirror true sampling variability.

– Table 1 about here –

In all, these results suggest that time-level clustering is important to consider in datasets where both individual and cluster level observations are present. Both simple MLM and OLS models as well as those that include a lagged cluster mean (OLS-LDV and MLM-LDV) perform poorly when long memory is ignored. As d increases these models produce biased parameter estimates, with less precision, and standard errors that are too small. On the other hand, MLM-ARFIMA and OLS-ARFIMA produce unbiased parameter estimates when fractional integration is considered. Yet, these two choices are not comparable with respect to their standard errors – OLS estimates of level-2 variables’ standard errors will be too small, thus elevating the risk of Type I error. Thus, we advocate the MLM-ARFIMA model when encountering data with a dynamic

¹⁹ We present only the optimism for the between-day effects. The optimism estimates hover near 100 for the within-day effects in all six models.

component across cross-sections. In the next section we further compare these methods in the context of the 2008 political campaign, exploring the factors that influenced candidate evaluation.

Application: Dynamic Processes and the Political Campaign

Both scholarly and convention wisdom suggest that as dire economic circumstances became a major focus of the campaign, the 2008 presidential election increasingly favored Barack Obama (for a thorough review, see Kenski, Hardy and Jamieson 2010). In this application, we empirically assess this argument, exploring the extent to which real economic conditions and perceived economic evaluations shaped evaluations of Senators Barack Obama and John McCain.

The National Annenberg Election Survey (NAES) affords unique leverage in addressing this issue, as it includes daily interviews from early January to November 3, 2008. The average number of respondents interviewed each day was $n=183$ (range = [40, 383]). The data allow us to simultaneously examine how stable, unchanging individual-level factors and dynamic, campaign-specific processes influence political behaviors and judgment. Recent work, for example, has demonstrated that assessments of the parties, candidates, and issues change - in some cases dramatically - throughout the course of the campaign (Kenski, Hardy, and Jamieson 2010; Brady and Johnston 2006; Stokes, Campbell, and Miller, 1958; Campbell, Converse, Miller, and Stokes 1960). Using the same models from our simulations, we test these approaches to account for individual and aggregate effects simultaneously.

Our dependent variable was constructed by subtracting positive Obama evaluations from positive McCain evaluations. We first constructed an evaluation scale separately for each candidate from five survey questions - [Obama/McCain] is a strong leader, [Obama/McCain] is trustworthy, [Obama/McCain] has experience to be president, [Obama/McCain] is ready to be president, and favorability toward [Obama/McCain] (Obama, $\alpha=0.96$; McCain, $\alpha=0.93$).

Each question was asked on a 0 to 10 scale, with 10 indicating a more positive evaluation. The entire sample average for evaluations of McCain was $M=6.10$, $SD=2.42$; for Obama, $M=5.59$, $SD=2.87$. We then subtracted evaluations of McCain from evaluations of Obama to obtain a relative evaluation scale. As such, 0 indicates equal evaluations of the candidates, positive scores indicate a more positive evaluation of Obama relative to McCain, and negative scores indicate a more positive evaluation of McCain relative to Obama.

We include several covariates in our analysis. Economic evaluations were assessed with a single item measuring the extent to which respondents believe the country's economic conditions are better than a year ago. It is a five point scale where higher scores indicate a positive evaluation ($M=3.46$, $SD=1.18$). Party identification ranges from 1 to 7 with higher scores denoting Democratic identification ($M=4.22$, $SD=2.21$). We also include age (in years), gender (1=Female, 0=Male), and the natural logarithm of income ($M=3.81$, $SD=0.86$). Finally, we merged the daily Dow Jones Industrial Average (DJIA) with our data, recoded so that a unit change corresponds to a 100 point change.²⁰ This allows us to test whether changes in real economic factors influence candidate evaluation – something that wouldn't be possible to test using a single cross-sectional data set.

As was the case in the simulations, we separate individual/within day and aggregate/between day effects. In this example, the distinction could be quite important. It is possible that, at the aggregate, voters respond to changes in economic conditions and that perceptions of the economy drive day-level changes in candidate evaluation; yet, it is uncertain whether this effect is equivalent at the individual-level. Moreover, failing to account for the dynamic nature of between-day effects could lead to incorrect parameter estimates and erroneous conclusions about the economy and candidate evaluation. We anticipate that autocorrelation will

²⁰ The NAES interviews respondents every day of the week. To avoid dropping weekend observations from our data, we use the previous Friday's DJIA value for observations on Saturday and Sunday.

be a concern for the level-2 model, in that our aggregated variables at t should be strongly related to themselves at $t-1, t-2...t-k$. As we demonstrate in our simulations, only the MLM-ARFIMA models should effectively resolve the problem of serial correlation.

To prepare our data, we calculated the day-level means for comparative candidate evaluation, personal income, economic evaluations, and party identification. We then estimated ARFIMA models for each variable in order to generate a white noise series for each (i.e., \bar{X}_t^* and \bar{Y}_t^*).²¹ This initial filter was used for estimates of the OLS-ARFIMA and MLM-ARFIMA models. Congruent with the simulations, the second filter involved subtracting out individual observations from the day level means for both x and y . Thus, for our preferred method, we are implementing our double-filtering technique – first, we obtain the non-deterministic day level means; second, we filter the individual level data through the day level means.

– Table 2 about here –

Table 2 presents the results for eight modeling strategies. As in the simulations, we expect that failing to account for clustering, and running an OLS or MLM model that assumes errors are independent, will lead to distorted results. Insofar as a dynamic component accounts for much of the autocorrelation in errors, the most accurate model should be the MLM-ARFIMA model. This is essentially what we find with noticeable differences between the models in the substantive conclusions that would be reached.

The largest differences in these models can be found in the between-day portion of the model, since this is the component of the model most affected by autocorrelation. Consider the

²¹ For our dependent variable, diagnostic tests indicated that we could not reject the null that $d=0$ (KPSS=0.29, $p>0.1$), and we can reject the null that $d=1$ (Dickey-Fuller $Z=-251.73$, $p<0.01$). We do find evidence of a non-zero fractional difference parameter ($d=0.2$) and estimate a (0, 0.2, 0) noise model. For our other variables, we cannot reject the unit root hypothesis for the DJIA and find that a (1, 0, |7|) model best fits the data. Personal income was found to follow a (0, 0, 0) process (KPSS=-6.82, $p<0.01$, Dickey-Fuller $Z=-294.92$, $d=0.03$); PID followed a (0, 0, 0) process (KPSS=0.07, $p>0.10$, Dickey-Fuller $Z=-268.51$, $p<0.01$, $d=0.08$); and economic evaluations a (1, 0.45, |6,7|) process (KPSS=2.79, $p<0.01$, Dickey-Fuller $Z=-150.59$, $d=0.45$).

size of the standard errors in the “between-day” effects in Table 2. The standard errors for the between effects are significantly smaller for all models relative to the MLM-ARFIMA model. However, the changes in the standard errors are relatively modest – which is likely due to the fact that d was not particularly large for any of these variables (with the exception of the DJIA).²² Likewise, the parameter estimates for the between-effects” are also quite different when we fail to account for autocorrelation in the level-2 variables. With the exception of OLS-ARFIMA and MLM-ARFIMA, there is a significant relationship between candidate evaluation and the DJIA, where economic strength relates to a more positive evaluation of McCain. However, after accounting for the ARFIMA process in both candidate evaluations and the DJIA, this relationship disappears. Both the OLS-ARFIMA and the MLM-ARFIMA models show a positive, albeit negligible and statistically non-significant, relationship between real economic conditions and candidate evaluation.

We find a similar pattern for aggregate economic evaluations. Failure to account for the ARFIMA process in economic evaluations leads to an erroneous conclusion that greater confidence in the economy translates to a more positive assessment of McCain. However, this effect also disappears in the OLS-ARFIMA and MLM-ARFIMA models, suggesting that the autocorrelation in the level-2 residuals accounts for this effect. In fact, the only variable to remain statistically significant in the between-effects, regardless of the method used, is PID. Increased Democratic Party identification in the electorate translates to an increasingly positive evaluation of Obama.

There are also important differences contrasting the within-effects to the between-effects in our models. The OLS-ARFIMA and MLM-ARFIMA models indicate that, while economic

²² Why is d so low for our dependent variable, especially in comparison with values closer to 0.7 and 0.8 found in many monthly time series (e.g. Box-Steffensmeier and Smith 1996, 1998; Lebo, Walker, and Clarke 2000)? The biggest factor is the noisiness of the small sample daily data. A great deal of movement in the series is simply due to measurement error in the daily aggregates. This added randomness drives down the level of memory and gives lower levels of d than would be found in most RCS time series. Thus, these models may greatly *understate* the differences between these modeling approaches than would be found using, say, 200 months of Gallup data with 1,000 respondents each month giving more accurate estimates of the true population values.

evaluations do not exert a striking effect at the aggregate, they are a significant predictor of candidate evaluation at the individual level. In other words, perceptions of economic conditions are related to candidate evaluation. Voters who believe the economy is better off than a year ago are more supportive of McCain; those who view the economy as worse than a year ago are more supportive of Obama. However, changes in aggregate economic conditions do not relate to changes in candidate evaluation. We believe this finding underscores the importance of separating within and between cluster effects (see also Bafumi and Gelman 2007; Bartels 2009a), or aggregate versus individual effects (Green and Shapiro 1994), and failing to separate individual from aggregate effects increases the risk of the ecological fallacy (Kramer 1983; King, Tanner, and Rosen 2004).

It is also important to point out that, unlike the simulations, we find what appears to be negligible differences with respect to the OLS-ARFIMA versus MLM-ARFIMA models. This is largely due to the fact that there is lessened degree of pooling in the data, with an intra-class correlation less than 0.01 - again a function of using daily data. This leads to substantially less heterogeneity in intercepts, which explains the negligible differences. Nevertheless, a Likelihood-Ratio test indicates that the level-2 intercepts are non-zero and the MLM-FI is preferred to the OLS-FI model ($\lambda^2[1] = 33.69$, adjusted p-value < 0.01, [Verbeke and Molenberghs 2000]).

Congruent with our simulations, Table 2 illustrates several important points: First, one should separate individual and aggregate effects, and it is important to account for the ARFIMA processes in both the dependent variable and covariates. Failing to do so heightens the chance of reaching erroneous conclusions about day-level processes. In this example, we find that the relationship between aggregate economic considerations and candidate evaluation may be spurious. Accounting for the ARFIMA process resulted in different conclusions regarding the

relationship between the economy, economic evaluations, and candidate evaluation – namely, there is a negligible relationship between changes in the economy and economic confidence and candidate evaluation. Second, we believe it important to underscore that our findings do not mean evaluations of the economy were unimportant in 2008. On the contrary: the within effects demonstrate that voters who felt the economy was in better shape were more likely to favor McCain to Obama. While aggregate changes in the economy and economic evaluations did not manifest in marginal increases for either candidate, at the individual level, we find a relatively robust relationship between beliefs about the economy and candidate evaluation. The substantive take-home point of this is that the economy matters to individual-level evaluations but that the downturn in the stock market and in economic judgments did not help overall evaluations of Senator Obama. In the next section, we elaborate on this point, extending the MLM model to explore changes in the within-day effects over the course of the campaign.

Modeling Dynamic Day Level Effects

Unlike OLS, one of the advantages of the MLM is that one may specify that the effects of covariates also randomly vary. To examine this in the NAES, we estimated a series of MLM-ARFIMA models, allowing the within-effects slopes to vary across days. Specifically, we estimated a model that allows the intercept for both economic evaluations and PID to vary across days.²³ From this model, we retrieved the day-level slopes using Empirical Bayes estimates (Skrondal and Rabe-Hesketh 2004). These values are plotted in Figure 6, with a Loess smoother to illustrate the trends of PID and economic evaluations over time. The solid line represents the fixed effect of PID and economic evaluations.

²³ We compared the random intercept to the random intercept plus random slope models (random intercept relative to random intercept+random economic evaluations model, $\lambda^2[1] = 8.26$, adjusted p-value<0.01; random intercept relative to random intercept+random PID model, $\lambda^2[1] = 963.00$ adjusted p-value<0.01)

– Figure 6 about here –

In line with the notion that campaigns activate and reinforce latent predispositions (Lazarsfeld, Berelson, and Gaudet 1944), we find a marked increase in the relationship between PID and candidate evaluation over the course of the campaign. The slopes are nearly twice as large at the end of the campaign relative to the beginning, suggesting the campaign helps voters connect their partisan beliefs to the candidates. Likewise, the bottom panel in Figure 6 demonstrates that the relationship between economic considerations and candidate evaluation grows stronger over the course the campaign. Near the end of the campaign, the estimated slope for economic considerations is nearly twice as large as what it was at the beginning of the campaign. Aside from these substantive implications, the clear point here is the added value of using a multi-level model to investigate political relationships over time.

Concluding Remarks

What political methodologists already knew about the problems of time series data, coupled with our investigations here, indicate that time-level clustering is an important issue to consider in data-sets that follow an RCS design. We believe that much of the prior work using RCS data has been unsatisfactory – analyzing either within-day processes or between-day processes exclusively, not both simultaneously. In our simulation analysis, we demonstrate that failing to specify dynamic effects, when dynamic, between-day processes exist, can lead to biased parameter estimates and incorrectly estimated standard errors. We demonstrate that the preferred way to deal with this problem is a two-step filtering process, where means are retrieved and a level-2 ARFIMA model is first specified followed by filtering level-1 data through these estimates. The MLM-ARFIMA model had desirable properties relative to the other seven tested models. Every one

of the other models encounters problems in one area or another but MLM-ARFIMA always performs best, especially as memory in the aggregates gets longer.

It is worth noting the various points of flexibility within our framework. Depending upon the length of T , one might estimate an ARMA, ARIMA, or ARFIMA model to create an appropriate noise model. For example, an RCS design consisting of 20 consecutive NES surveys would likely have some autocorrelation at level-2 but a noise model would be best chosen among simpler ARMA models. And, if the best model was simply $(0,0,0)$, that is, no autocorrelation existed in the aggregate, then the model reduces to simple mean-centering of the level-1 units (as suggested by Bafumi and Gelman 2007). On the other hand, if one were to reanalyze the 87 quarters of data used by Box-Steffensmeier et al. (2004), it would be best to begin with an ARFIMA noise model at level-2 similar to what those researchers used studying the data strictly as time series.

In addition, as demonstrated in our NAES example, the model can include time varying coefficients for some covariates. To be sure, with so much data, the RCS design is a great resource for studying time-varying relationships. Allowing the constant and coefficients to vary from one wave to the next while also measuring level-2 factors, means that the effects of level-1 variables can be seen to rise and fall according to the daily context. Still, without a method such as double filtering, the inferences from such an exercise would be suspect.

Our MLM-ARFIMA framework can certainly be extended to PCSTS designs, but with two notable caveats. First, the number of pseudo-waves that can be compiled into a RCS design may be quite high, perhaps running into the hundreds of consecutive data sets. With PCSTS, however, datasets are rarely very long. Yearly data by country often tops out at $t=65$ for the post-war era. True panels of individual-level data are unlikely to ever approach the t of an RCS design. One can only wish for something like the three- and four-wave panels sometimes seen in the National

Election Study or the British Election Study to be carried on at frequent intervals for decades with the same individuals. So, with a shorter t , the PCSTS analyst will likely need to estimate a simpler noise model at the aggregate-level. ARMA and ARIMA methods are just particular types of ARFIMA models where d equals 0 and 1, respectively. But ARMA and ARIMA are more dependable to estimate up until the point of about $t = 50$ (Dickinson and Lebo 2007).

Second, the PCSTS analyst will have several other methods at their disposal that will help alleviate problems of autocorrelation. Panel corrected standard errors (Beck and Katz 1995) are one such tool. Differencing or using a lagged dependent variable are two imperfect solutions but they are improvements on simply pooling the data or ignoring the sequence of the waves. The use of dynamic panels (Baltagi 2005) is also an elegant solution for PCSTS designs. By differencing all the variables, a dynamic panel allows unit-specific idiosyncrasies to cancel each other out. Our double filtering method can be added to this list of solutions but, admittedly, has a lot more competition in that particular toolbox.

Considering our findings, we encourage researchers to adopt the MLM-ARFIMA model when analyzing RCS data. By using multilevel models to study RCS data, not only can researchers capture contemporaneous variation, but they can also directly model dynamic processes. Taken together, these results suggest that time-level clustering is important to consider, and with greater attention to simultaneously modeling static and dynamic processes, this will provide a richer depiction of various political and social phenomena.

References

- Achen, Christopher. 2000. "Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables." Presented at the annual meeting of the APSA.
- Athey, Susan and Guido Imbens. 2006. "Identification and Inference in Nonlinear Difference in Differences Models." *Econometrica* 74: 431-497.
- Bafumi, Joseph. and Andrew Gelman. 2006. "Fitting Multilevel Models When Predictors and Group Effects Correlate." Presented at the Annual Meeting of the MPSA, Chicago, IL.
- Baltagi, Badi. 2005. *Econometric Analysis of Panel Data*. West Sussex, England: Wiley.
- Bartels, Brandon. 2009a. "Beyond Fixed versus Random Effects: A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel Data." The Society for Political Methodology Working Paper.
- Bartels, Brandon. 2009b. "The Constraining Capacity of Legal Doctrine on the US Supreme Court." *American Political Science Review* 103:474-495.
- Bates, Douglas and Martin Maechler. 2010. "lme4: Linear mixed effects models using Eigen and Eigen++." R package version 0.999375-37. <http://CRAN.R-project.org/package=lme4>.
- Beck, Nathaniel and Jonathan Katz. 1995. "What to Do (And Not to Do) With Time-Series Cross-Section Data." *American Political Science Review* pp. 634-647.
- Beck, Nathaniel and Jonathan Katz. 2007. "Random Coefficient Models for Time-Series- Cross-Section Data: Monte Carlo Experiments." *Political Analysis* 15:182.
- Box, George and Gwilym Jenkins. 1976. *Time Series Analysis, Forecasting, and Control*. San Francisco: Holden Day.
- Box-Steffensmeier, Janet M., Suzanna De Boef & Tse-min Lin. 2004. "The Dynamics of the Partisan Gender Gap." *The American Political Science Review* 98:515-528.
- Box-Steffensmeier, Janet M., and Renee M. Smith. 1996. "The Dynamics of Aggregate Partisanship." *American Political Science Review* 90: 567-80.
- Brady, Henry and Richard Johnson. 2006. *Capturing Campaign Effects*. Ann Arbor, MI: University of Michigan Press.
- Brooks, Deborah Jordan and John G. Geer. 2007. "Beyond Negativity: The Effects of Incivility on the Electorate." *American Journal of Political Science* 51:1-16.
- Bryk, Anthony and Stephen Bryk. 1992. *Hierarchical Linear Models in Social and Behavioral Research*. Newbury Park: Sage.

- Brown, D. S. and A. M. Mobarak. 2009. "The Transforming Power of Democracy: Regime Type and the Distribution of Electricity." *American Political Science Review* 103:193– 213.
- Campbell, Angus, Philip E. Converse, Warren E. Miller and Donald E. Stokes. 1960. *The American Voter*. New York: Wiley and Sons.
- Canes-Wrone, Brandice, David W. Brady and John F. Cogan. 2002. "Out of Step, out of Office: Electoral Accountability and House Members' Voting." *APSR* 96:127–140.
- Clarke, Harold, Stewart, Marianne, Ault, Mike, and Euel Elliott. 2005. "Men, Women and The Dynamics of Presidential Approval." *British Journal of Political Science* 35: 31-51.
- Clarke, Harold and Matthew Lebo. 2003. "Fractional (Co)integration and Governing Party Support in Britain." *British Journal of Political Science* 33: 283-301.
- Dickinson, Matthew J. and Matthew J. Lebo. 2007. "Reexamining the Growth of the Institutional Presidency, 1940-2000." *Journal of Politics* 69: 206-19.
- DeBoef, Suzanna and Paul M. Kellstedt. 2004. "The Political (And Economic) Origins of Consumer Confidence." *American Journal of Political Science* 48:633–649.
- DiPrete, Thomas and David Grusky. 1990. "The Multilevel Analysis of Trends with Repeated Cross-Sectional Data." *Sociological Methodology* 20: 337-368.
- Enders, Walter. 2004. *Applied Econometric Time Series*. New York: Wiley and Sons.
- Freedman, Paul, Michael Franz and Kenneth Goldstein. 2004. "Campaign Advertising and Democratic Citizenship." *American Journal of Political Science* 48:723–741.
- Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Gelman, Andrew, David Park, Boris Shor, Joseph Bafumi, and Jeronimo Cortina. 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way they Do*. Princeton, NJ: Princeton University Press.
- Gidengil, Elizabeth and Agnieszka Dobrynska. 2003. "Using a Rolling Cross Section Design to Model Media Effects: The Case of Leader Evaluations in the 1997 Canadian Election." Paper presented at the annual meeting of the American Political Science Association
- Granger, Clive and Paul Newbold. 1974. "Spurious Regression in Econometrics." *Journal of Econometrics* 2: 111-120.
- Green, Donald and Ian Shapiro. 1994. *Pathologies of Rational Choice: A Critique of Applications in Political Science*. New Haven: Yale University Press.
- Hamilton, James. 1994. *Time Series Analysis*. Princeton, NY: Princeton University Press.

- Heckman, James and Brook Payner. 1989. "Determining the Impact of Federal Antidiscrimination Policy on Economic Status of Blacks: A Study of South Carolina." *The American Economic Review* 79: 138-177.
- Honaker, James and Gary King. 2010. "What to Do about Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54: 561-581.
- Hurst, H. Edwin. 1951. "Long-term Storage Capacity of Reservoirs." *Transactions of the American Society of Civil Engineers* 116: 770-799.
- Jerit, Jennifer, Jason Barabas and Toby Bolsen. 2006. "Citizens, Knowledge, and the Information Environment." *American Journal of Political Science* 50(1):266-282.
- Johnston, Richard, Michael Hagen and Kathleen Hall Jamieson. 2004. *The 2004 Presidential Election and the Foundation of Party Politics*. Cambridge: Cambridge University Press.
- Keele, Luke and Nathan Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14:186-205.
- Kenny, Christopher and Michael McBurnett. 1992. "A Dynamic Model of the Effect of Campaign Spending on Congressional Vote Choice." *American Journal of Political Science* pp. 923-937.
- Kenski, Kate, Bruce Hardy, and Kathleen Hall Jamieson. 2010. *The Obama Victory: How Media, Money, and Message Shaped the 2008 Election*. New York: Oxford University Press.
- King, Gary, M.A. Tanner and O. Rosen. 2004. *Ecological inference: New Methodological Strategies*. Cambridge University Press.
- Kramer, G. H. 1983. "The Ecological Fallacy Revisited: Aggregate-Level Versus Individual-Level Findings on Economics and Elections, and Sociotropic Voting." *American Political Science Review* 77:92-111.
- Lau, Richard, D.J. Andersen and David Redlawsk. 2008. "An Exploration of Correct Voting in Recent US Presidential Elections." *American Journal of Political Science*
- Lazarsfeld, Paul, Berelson, Bernard, and Hazel Gaudet. 1944. *The People's Choice: How The Voters Makes up his Mind in a Presidential Campaign*. New York: Duell.
- Lebo, Matthew, Robert Walker, and Harold Clarke. 2000. "You Must Remember This: Dealing with Long Memory in Political Analyses." *Electoral Studies* 19:31-48.
- Lebo, Matthew, Adam McGlynn and Gregory Koger. 2007. "Strategic Party Government: Party Influence in Congress, 1789-2000." *American Journal of Political Science* 51:464-481.
- MacKuen, Michael, Erickson, Robert and James Stimson, 1989. "Political Parties, Public Opinion, and State Policy." *American Political Science Review* 83: 729-750.

- MacKuen, Michael B., Robert S. Erikson and James A. Stimson. 1992. "Peasants or Bankers? The American Electorate and the U.S. Economy." *The American Political Science Review* 86:597–611.
- Mishler, William and Reginald Sheehan. 1996. "Public opinion, the attitudinal model, and Supreme Court decision making: A micro-analytic perspective." *Journal of Politics*, 169–200.
- Moy, Patricia, Michael Xenos and V.K. Hess. 2006. "Priming effects of late-night comedy." *International Journal of Public Opinion Research* 18:198.
- Raudenbush, Stephen and Anthony Bryk. 2002. *Hierarchical Linear Models*. Newbury Park, CA: Sage.
- Romer, David. 2006. *Capturing campaign dynamics, 2000 and 2004: The National Annenberg Election Survey*. University of Pennsylvania Press.
- Shayo, Moses. 2009. "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution." *American Political Science Review* 103:147–74.
- Shor, Boris, Joseph Bafumi Keele Luke and David Park. 2007. "A Bayesian Multilevel Modelling Approach to Time Series Cross-Sectional Data." *Political Analysis*.
- Skrondal, Anders and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.
- Snijders, Tom and Roel Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Steenbergen, Marco R. and Bradford S. Jones. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46:218–37.
- Stoker, Laura and M.Kent Jennings. 2008. "Of Time and the Development of Partisan Polarization." *American Journal of Political Science* 52:619–635.
- Stokes, Donald, Campbell, Angus, and Warren Miller. 1958. "Components of Electoral Decision." *The American Political Science Review* 52: 367-387.
- Stroud, Natalie. 2008. "Media use and political predispositions: Revisiting the concept of selective exposure." *Political Behavior* 30:341–366.
- Verbeke, Geert and Geert Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Voeten, Erik. 2008. "The Impartiality of International Judges: Evidence from the European Court of Human Rights." *American Political Science Review* 4:417–433.
- Wooldridge, Jeffrey. 2001. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Figure 1: Coefficient Estimates in Two Naïve Models

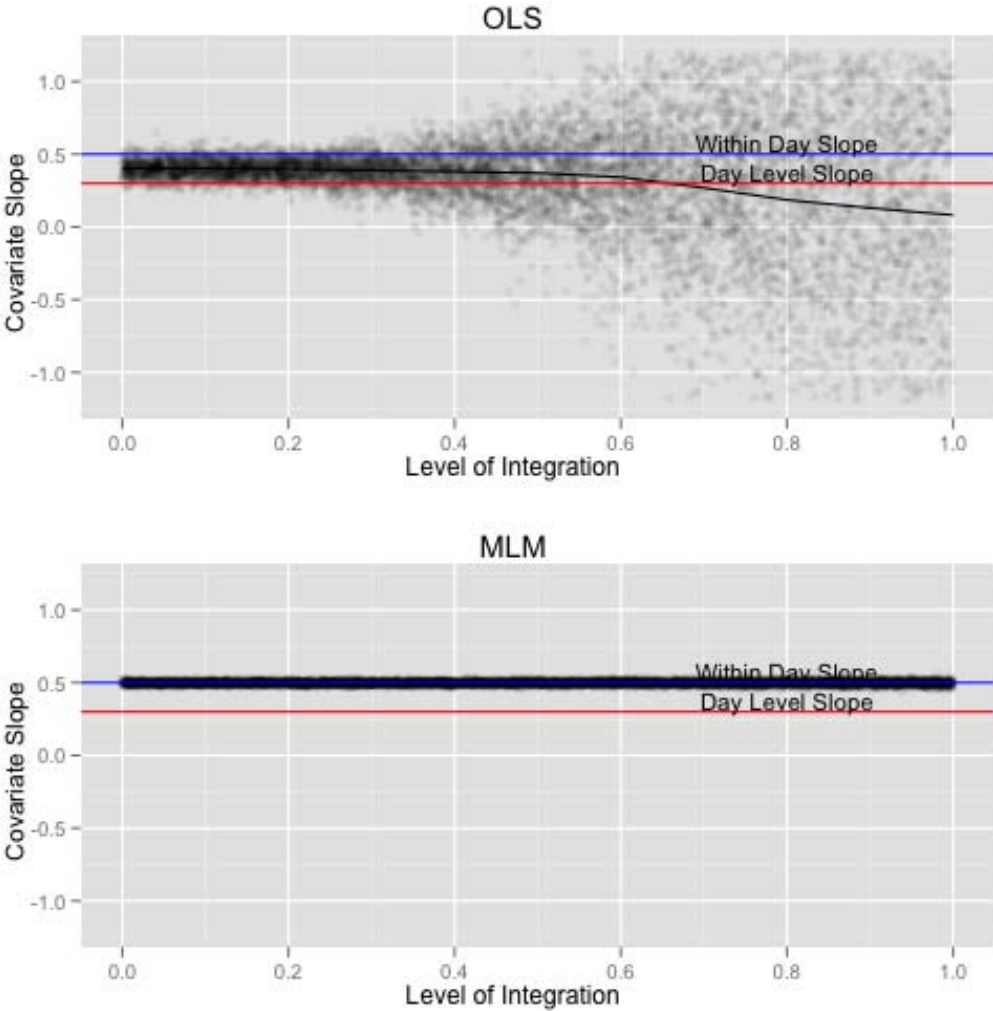
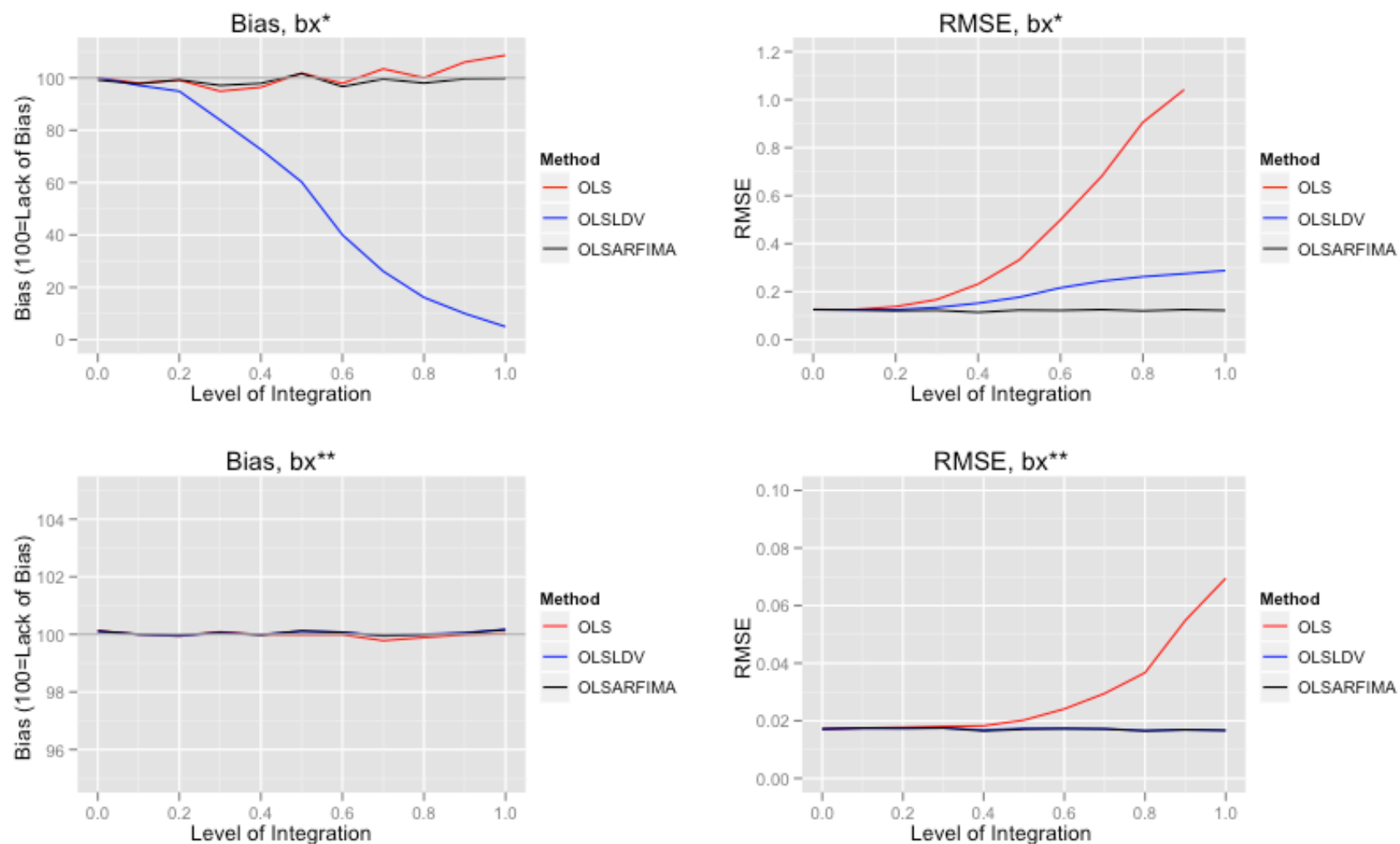
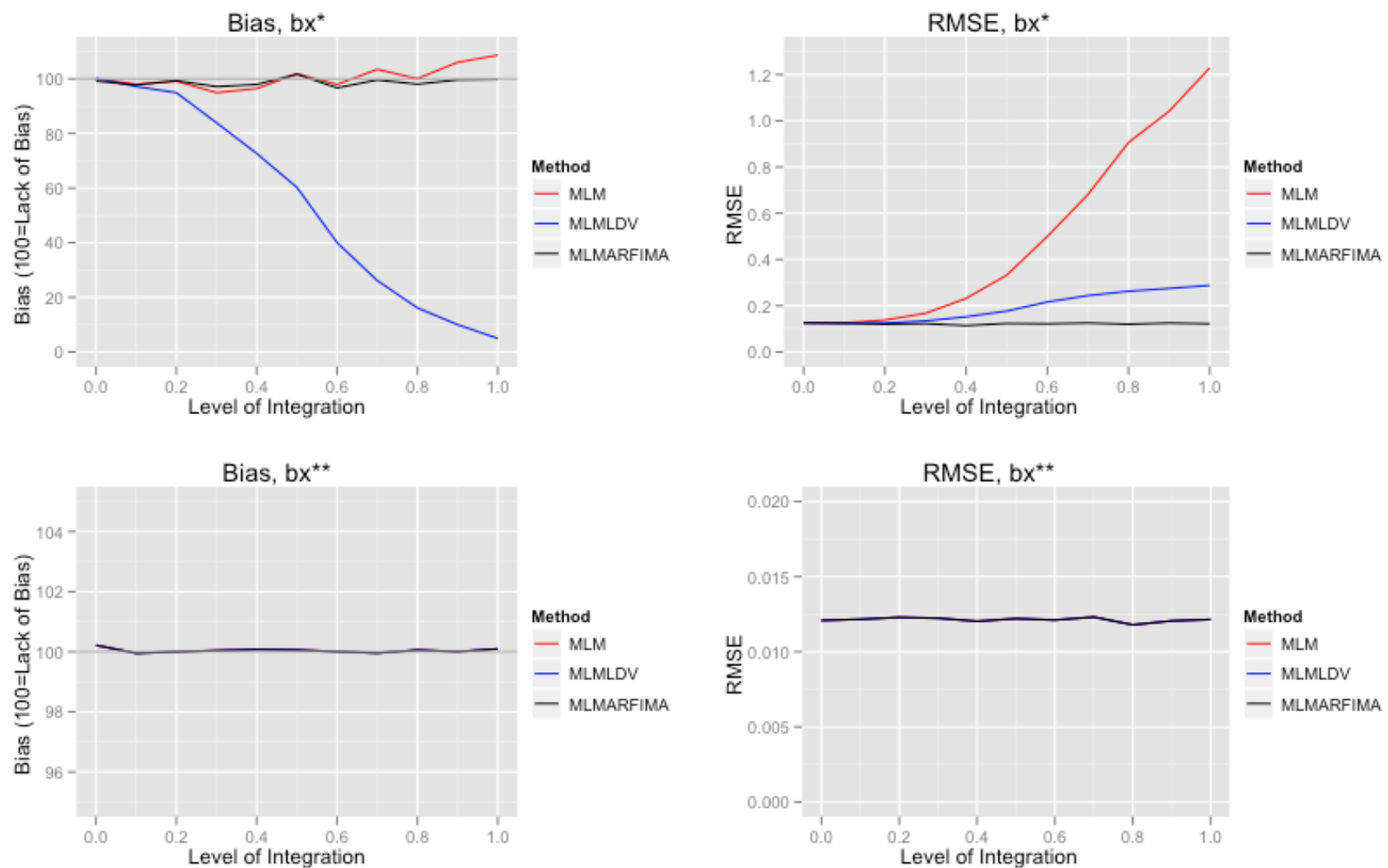


Figure 2: Bias and RMSE for OLS Coefficients*



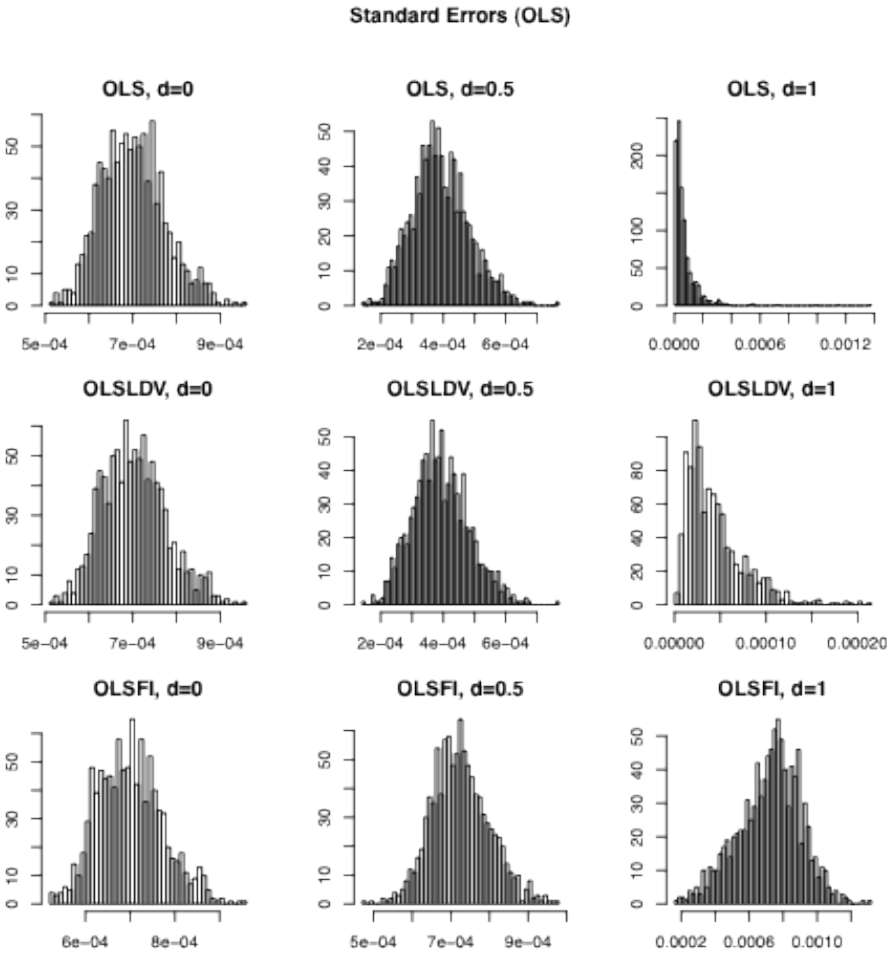
* For the OLS and OLS-LDV models this is the coefficient for \bar{X}_t . For the OLS-ARFIMA models it is the coefficient for \bar{X}_t^* . Lines in the bottom panel overlap with the OLS-LDV line in the bottom-right panel hidden by the OLS-ARFIMA line.

Figure 3: Bias and RMSE for Multilevel Models*



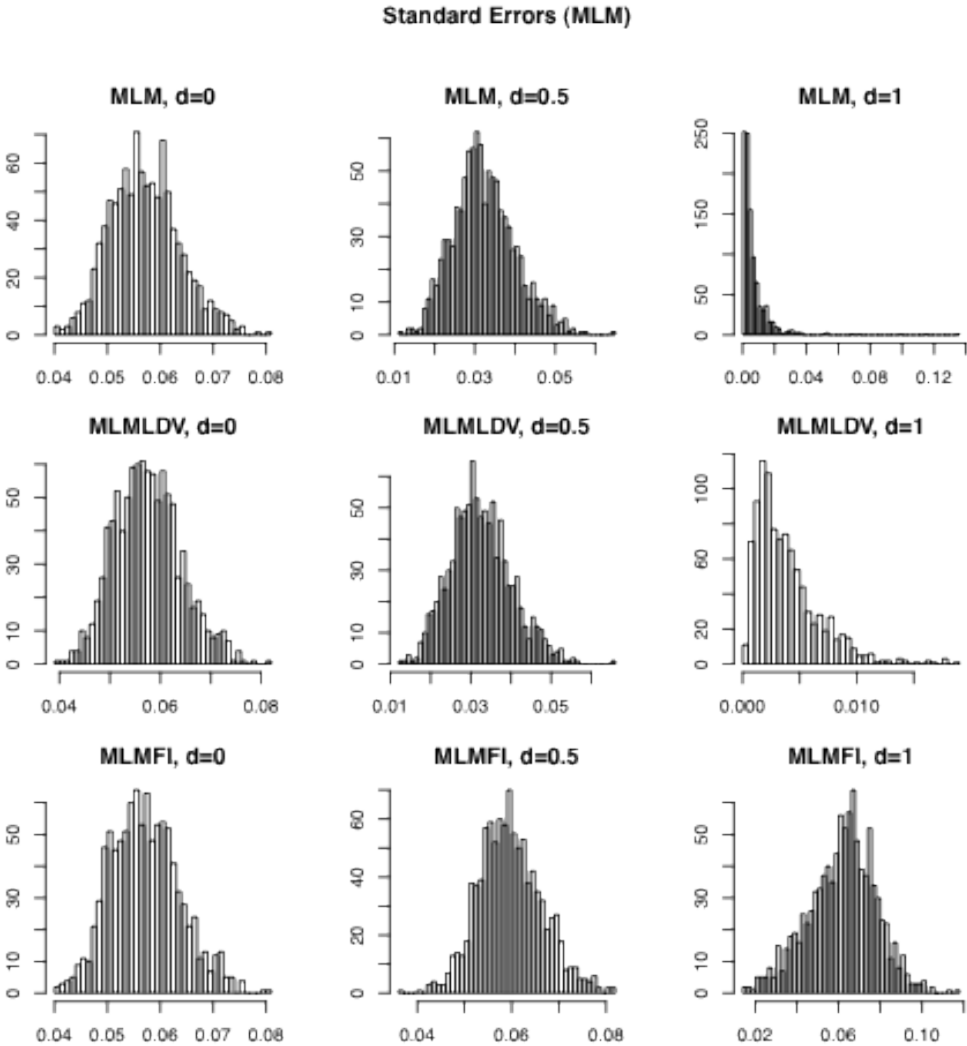
* For the MLM and MLM-LDV models this is the coefficient for \bar{X}_t . For the MLM-ARFIMA models it is the coefficient for \bar{X}_t^* . All three lines overlap in the bottom panels.

Figure 4: Standard Errors for βx^* , OLS Models*



For the OLS and OLS-LDV models this is the standard error of the coefficient for \bar{X}_t . For the OLS-ARFIMA models it is the standard error of the coefficient for \bar{X}_t^* .

Figure 5: Standard Errors for βx^* , Multilevel Models*



For the MLM and MLM-LDV models this is the standard error for the coefficient for \bar{X}_t . For the MLM-ARFIMA models it is the standard error of the coefficient for \bar{X}_t^* .

Figure 6: Varying Impact of Economic Evaluations and PID on Candidate Evaluations, 2008 Campaign

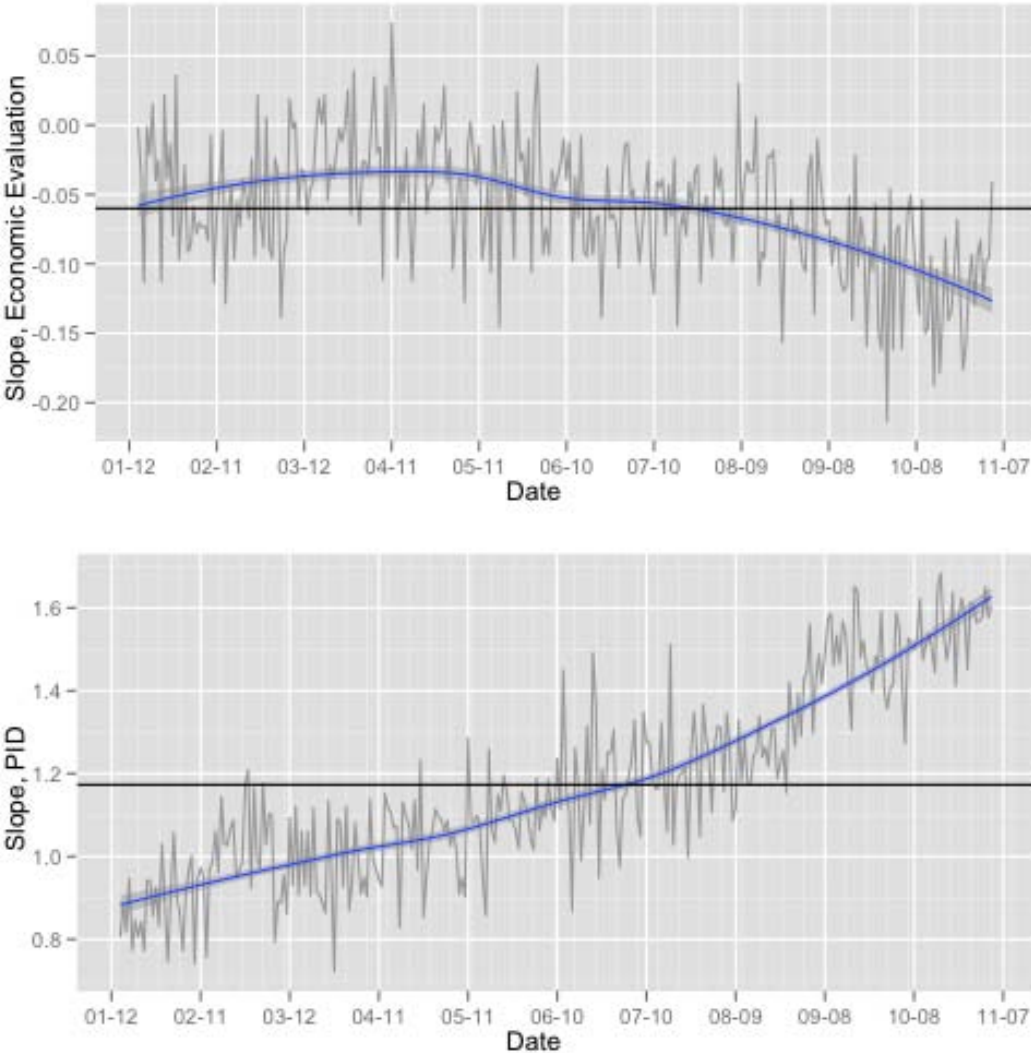


Table 1: Optimism Index for Six Modeling Approaches

d	Between Day Effects (bx^*)					
	<i>OLS</i>	<i>OLSLDV</i>	<i>OLS-ARFIMA</i>	<i>MLM</i>	<i>MLM-LDV</i>	<i>MLM-ARFIMA</i>
0	747	1818	734	105	255	103
0.1	747	1695	723	104	238	101
0.2	818	1601	702	113	224	99
0.3	1025	1346	713	138	188	100
0.4	1462	1101	666	190	153	94
0.5	2189	1007	721	269	139	101
0.6	3432	1419	709	396	196	100
0.7	4875	2283	730	534	316	103
0.8	6525	3636	702	685	506	99
0.9	7480	5172	732	766	723	103
1.0	8344	7231	711	843	1011	100

For the OLS, OLS-LDV, MLM, and MLM-LDV models these are based on the standard errors of coefficients for \bar{X}_t . For the OLS-ARFIMA and MLM-ARFIMA models they are based on the standard error of the coefficient for \bar{X}_t^* .

Table 2: Six Models of Candidate Evaluation

	<i>OLS-Naive</i>	<i>OLS</i>	<i>OLS-LDV</i>	<i>OLS-FI</i>	<i>MLM-Naive</i>	<i>MLM</i>	<i>MLM-LDV</i>	<i>MLM-FI</i>
Between Effects								
Economic Evaluation	---	-0.334 (.1049)	-0.288 (.1052)	0.057 (.158)	---	-0.330 (.1340)	-0.285 (.1300)	0.050 (.190)
PID (Democrat)	---	1.312 (.0972)	1.240 (.0979)	1.256 (.0980)	---	1.277 (.1231)	1.212 (.1200)	1.233 (.1171)
Personal Income	---	0.032 (.2127)	0.110 (.2131)	0.120 (.208)	---	-0.017 (.270)	0.068 (.2618)	0.090 (.2491)
DJIA	-0.004 (.0014)	-0.006 (.0016)	-0.005 (.0016)	0.008 (.008)	-0.004 (.0014)	-0.006 (.0021)	-0.005 (.0020)	0.008 (.01)
Lag Y			0.242 (.0417)				0.236 (.0520)	
Intercept	-4.373 (.1929)	-4.022 (1.031)	-4.200 (1.031)	-5.42 (.0899)	-4.373 (.1929)	-3.670 (1.313)	-3.910 (1.272)	-5.218 (1.076)
Within Effects								
Economic Evaluation	-0.069 (.0139)	-0.064 (.014)	-0.064 (.0140)	-0.064 (.0140)	-0.067 (.0139)	-0.064 (.014)	-0.064 (.0140)	-0.064 (.0140)
PID (Democrat)	1.200 (.0074)	1.200 (.0074)	1.200 (.0074)	1.200 (.0074)	1.200 (.0074)	1.200 (.0074)	1.200 (.0074)	1.200 (.0074)
Personal Income	-0.053 (.0196)	-0.055 (.0197)	-0.055 (.0197)	-0.055 (.0197)	-0.054 (.0196)	-0.055 (.0197)	-0.055 (.0197)	-0.055 (.0197)
Age	-0.008 (.001)	-0.008 (.001)	-0.008 (.001)	-0.008 (.001)	-0.008 (.001)	-0.008 (.001)	-0.008 (.001)	-0.008 (.001)
Female	0.245 (.033)	0.246 (.033)	0.246 (.033)	0.246 (.033)	0.245 (.033)	0.246 (.033)	0.246 (.033)	0.246 (.033)

Number of Days

291

N

42,100

Point estimates and standard errors (in parentheses). Dependent variable is candidate evaluation (Positive Evaluation of Obama – Positive Evaluation of McCain). Economic evaluation is coded such that high scores denote better economic conditions. Personal income is logged. Age is in years. DJIA=Dow Jones Industrial Average, which is recoded such that a unit increase corresponds to a 100 point change. Entries in bold indicate a coefficient two times the size of the standard error.