An Effective Approach to the Repeated Cross-Sectional Design

Matthew J. Lebo Stony Brook University
Christopher Weber University of Arizona

Repeated cross-sectional (RCS) designs are distinguishable from true panels and pooled cross-sectional time series (PCSTS) since cross-sectional units (e.g., individual survey respondents) appear but once in the data. This poses two serious challenges. First, as with PCSTS, autocorrelation threatens inferences. However, common solutions like differencing and using a lagged dependent variable are not possible with RCS since lags for i cannot be used. Second, although RCS designs contain information that allows both aggregate- and individual-level analyses, available methods—from pooled ordinary least squares to PCSTS to time series—force researchers to choose one level of analysis. The PCSTS tool kit does not provide an appropriate solution, and we offer one here: double filtering with ARFIMA methods to account for autocorrelation in longer RCS followed by the use of multilevel modeling to estimate both aggregate- and individual-level parameters simultaneously. We use Monte Carlo experiments and three applied examples to explore the advantages of our framework.

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 is not identical across all waves. The latter, what we call "pseudo-panels," have become increasingly common in political science and other social sciences.

Two types of pseudo-panel structures are distinguishable: the repeated cross-sectional (RCS) design and the "unbalanced panel." The latter has units that appear more than once, but not all cases appear in every time period.[1] Honaker and King (2010) discuss the unbalanced panel as a missing data problem and provide a multiple-imputation solution.[2] Yet, their solution cannot be applied where all but one observation of each case is missing in a data set comprised of multiple time-points. In RCS designs, this is exactly the problem. Whether it is a set of Gallup respondents nested within months, Supreme Court cases nested within court terms, or roll-call votes nested within congressional sessions, a unique set of cross-sectional units appears at each time-point.

In this article, we identify the statistical problems inherent in longer RCS designs and propose a method that negates autocorrelation problems. We estimate effects at two levels of analysis simultaneously and provide a reliable way to study time-varying relationships. Specifically, we use an autoregressive fractionally integrated moving average (ARFIMA) model to deal with autocorrelation at the aggregate level and then use a second filter, akin to mean centering in PCSTS designs, so that individual-level observations are also free of autocorrelation. Last, a multilevel model (MLM) estimates both individual- and aggregate-level effects while offering flexibility in

Matthew J. Lebo is Associate Professor, Department of Political Science, Stony Brook University, Stony Brook, NY 11794 (matthew.lebo@stonybrook.edu). Christopher Weber is Assistant Professor, School of Government and Public Policy, University of Arizona, Tucson, AZ 85721 (chrisweber@email.arizona.edu).

An earlier version of this article was presented at the 2010 annual meeting of the Southern Political Science Association. This article benefited greatly from help from Jason Barabas, Brandon Bartels, Johanna Dunaway, Stanley Feldman, Robert Franzese, Jim Garand, Taylor Grant, Bruce Hardy, Daniel Hopkins, Richard Johnston, Ellen Key, Gregory Koger, Maxwell Mak, Evan Parker-Stephen, Rosanne Scholl, and participants in department seminars at Louisiana State University and Stony Brook University. Replication materials and our supporting information are available online at http://thedata.harvard.edu/dvn/faces/study/StudyPage.xhtml?globalId=doi:10.7910/DVN1/22651&versionNumber=1. Please visit https://sites.google.com/a/stonybrook.edu/matthew-lebo/ for more information on obtaining and using a general R package to estimate ARFIMA-MLM models.

[1]See, e.g., Canes-Wrone, Brady, and Cogan (2002), Brown and Mobarak (2009), and Voeten (2008).

[2]This is especially useful for studies of international politics where cases like developing countries enter and leave the data set at different times or where other gaps appear in available data.

American Journal of Political Science, Vol. 59, No. 1, January 2015, Pp. 242–258

 DOI: 10.1111/ajps.12095

studying time-varying relationships. Our approach is not the first to propose using MLM for RCS data (see, e.g., Hopkins 2012; Lau, Anderson, and Redlawsk 2008), but it is the first to use a top-down approach. That is, we do not begin with microlevel data and apply fixes to the macro level. Instead, we use the techniques proven to be the most reliable for longer aggregate-level time-series analysis and then nest the individual-level data within the aggregates.

The article proceeds as follows: We first detail the properties of RCS data and discuss problems with available modeling techniques. Following that, we outline our ARFIMA-MLM method and present simulations that compare it to several alternatives. We then offer suggestions for applying our model to shorter RCS data and to true and unbalanced panels. Finally, we demonstrate our model in three separate examples using data from economic voting (Hopkins 2012), congressional behavior (Lebo, McGlynn, and Koger 2007), and the 2008 Annenberg Study (Kenski, Hardy, and Hall Jamieson 2010).

The Repeated Cross-Sectional Design

As a collection of individual-level data repeated at regular intervals, the RCS data structure can be extremely useful by adding a dynamic component to the study of cross-sectional units and by allowing the investigation of time-varying relationships.[3] RCS designs are increasingly prevalent in part due to the many ways in which they can be created.[4] The National Annenberg Election Study (NAES), for example, collected data on unique samples of voters for every day of three presidential election campaigns. Or one can create an RCS of over 300 months using CBS/NYT polls archived at the Interuniversity Consortium for Political and Social Research (ICPSR). The same can be done with hundreds of consecutive months of Gallup polls or Michigan's Survey of Consumers (e.g., Clarke, Stewart, Ault, and Elliott 2005; Hopkins 2012).[5]

And those are just some public opinion examples. Congressional roll calls nested within years since 1789 (Lebo, McGlynn, and Koger 2007), cases nested within Supreme Court terms since 1946 (Segal and Spaeth 2002), and public remarks by presidents nested within quarters since 1945 (Wood 2009) all fit the RCS format.[6] RCS data are widespread, and, as with longer pseudo- and true panels, the dynamic implications are rarely explored. Looking at just the 2010–13 issues of *American Political Science Review* and *American Journal of Political Science*, we find 42 articles where the underlying data structure—RCS, pseudo-panel, or true panel—could allow the use of our approach.[7]

Absent multilevel models, researchers have chosen to study RCS data in either the aggregate *or* the individual level. For the latter, some have simply pooled observations from all time-points (e.g., Jerit and Barabas 2012; Moy, Xenos, and Hess 2006; Romer 2006; Stroud 2008) or pooled subperiods of data (e.g., Blaydes and Chaney 2013 pool data from 700–1500 ACE by century). Pooling treats observations as if they were collected in a single cross-section. However, if units within time-points share unmeasured commonalities, standard errors may be incorrect. Or if one filters the time component via fixed effects to control for between-time-point effects, it limits the exploration to static processes and also assumes that parameter estimates pool around a common value. So pooling has its problems.

Alternately, one can skip individual-level analyses by collapsing data into mean values and applying time-series analyses to the aggregate data. Box-Steffensmeier, DeBoef, and Lin (2004), for example, study the gender gap using all available CBS/NYT surveys dating back to 1977 but aggregate respondents by quarter. In all, they use the responses of over 250,000 unique individuals, yet they analyze just 87 quarters of data (2004, 525).[8] Studying data

[3]In survey research, RCS designs also give researchers many benefits of traditional panel designs, but problems of attrition and response bias are avoided and sample sizes can be held steady.

[4]We do not use the terms *repeated cross-section* and *rolling cross-section* interchangeably. We take rolling to apply only to a survey design where all of the sample is identified at some initial time-point and the surveys are *rolled* out to sets of respondents at staggered dates. The Annenberg Study fits this description. Thus, a rolling cross-sectional design is one type of repeated cross-sectional design.

[5]Variables might not be measured in the same way across waves, and this can require corrections. Additionally, some RCS data sets may consist of surveys taken at irregular intervals. For example, Jerit, Barabas, and Bolsen (2006) pool dozens of Princeton Survey Research Associates surveys to study political knowledge over a 10-year period. But the frequency of the surveys is lumpy from year to year—some years have three surveys, some have none. The dynamic properties in such studies are beyond the scope of this article.

[6]For each of these, there should be concerns about autocorrelation at the aggregate level; party unity, Supreme Court liberalism, and presidential liberalism all exhibit time-dependent serial correlation.

[7]See Table S1 in the supporting information for a list that includes RCS structures and longer panels and pseudo-panels (e.g., the Polity IV data). Our approach may be useful where data are collected over a period that can or should be neatly segmented to account for dynamics over many time-points.

[8]Similar aggregation strategies have been used to study 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 1996). Johnston, Hagen, and Hall Jamieson (2004) study dynamic campaign effects by aggregating responses over multiple days of the

in the aggregate has theoretical support if causal ordering at the individual level is in question. For example, in the economic voting literature, Kramer (1983) argues that the state of the economy is an objective fact, and individuals' evaluations of it are either survey error or "partisanship, thinly disguised."[9]

If some independent variables vary only over time (e.g., the inflation rate) there is a natural tendency to construct a model in the aggregate. Yet, aggregating participants by day/week/month/quarter can reduce data sets to a thousandth of their original size. Indeed, 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. Since a multilevel model allows the use of *all* the data, the aggregate- versus individual-level debate is a false dichotomy. Researchers can explore complex relationships rather than entirely avoid an important level of analysis. In doing so, they can also investigate individual-level relationships that might vary over time.

Thus, two problems are evident. First, most published work using RCS has relied on techniques that study static *or* dynamic processes, not both. Second, ventures into MLM with RCS data have barely considered the statistical consequences of the various modeling choices. Given the wealth of RCS data available, we explore its challenges and consider the efficacy of several modeling choices.

Panels versus RCS: What's the Difference?

In a true panel design, N units are observed repeatedly over time, yielding an $N \times T$ data set and making autocorrelation likely in two directions. First, unit i at time t will be more correlated with unit j at time t than with unit j at other times. Second, the values for each unit i are likely correlated with each other over time. For example, in a Country by Year data set, regression errors are likely to be correlated within years as well as for each country. Importantly, neither type of autocorrelation disappears in an RCS design even though units are not repeated. First, dynamic autocorrelation remains. Memory over time, traceable through aggregates, likely exists between units more proximate to one another. If $\bar{Y}_t$ and $\bar{Y}_{t+1}$ are correlated, then $\varepsilon_{i,t}$ is correlated with $\varepsilon_{j,t+1}$ more than $\varepsilon_{i,t}$ is correlated with $\varepsilon_{k,t+2}$.[10] This holds since observations within each time-point are dispersed around a mean correlated with the mean of the adjacent time-point. That is, since $\bar{Y}_t = \mathrm{E}(y_{i,t})$ and corr ($\bar{Y}_t$ and $\bar{Y}_{t+1}$) $\neq 0$, then $\mathrm{corr}(\mathrm{E}(y_{i,t}), \mathrm{E}(y_{i,t+1})) \neq 0$ and $\mathrm{corr}(\mathrm{E}(\varepsilon_{i,t}), \mathrm{E}(\varepsilon_{i,t+1})) \neq 0$. Second, autocorrelated errors also exist due to day- (or month-, quarter-, or year-) specific effects.

How are these challenges to be handled? To begin, two common PCSTS approaches cannot or should not be used. Including a lagged dependent variable (LDV) is a popular way to handle problems of nonstationarity in PCSTS and traditional time series (Keele and Kelly 2006). A second alternative, looking at the differences in observations between time-points to render a random-walk series stationary, is also popular (Enders 2004). Yet, since each observation occurs but once in RCS data, these approaches cannot work. Values of $y_{i,t-1}$ are not available, so an LDV is not possible, nor is the use of a differenced dependent variable, $\Delta y_{i,t}$, created as $y_{i,t} - y_{i,t-1}$. A third unworkable approach, the use of panel-corrected standard errors (Beck and Katz 1995), is premised on observations repeating in every time-point (a true panel) and does not solve the potential bias in coefficients.[11]

As an RCS gets longer, the possibilities of modeling dynamic processes increase and should be pursued. 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. Next, we discuss fractional integration and outline the specifics of our ARFIMA-MLM approach.

ARFIMA Methods for Aggregate (Level-2) Data

The statistical properties of longer RCS data—once aggregated—have been well established. For example, provided enough time-points, monthly and quarterly public opinion data consistently prove to be fractionally integrated ("long-memory") when tested (Box-Steffensmeier, DeBoef, and Lin 2004; Box-Steffensmeier and Smith 1996, 1998; Byers, Davidson, and Peel 2000;

NAES while examining the individual-level data in separate models. Carey and Lebo (2006) use both levels of data in 70 consecutive months of British Gallup data to examine prospective versus retrospective voting models over a campaign cycle.

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

[10]Recent studies have consistently established that $\mathrm{corr}(\bar{Y}_t, \bar{Y}_{t+1}) \neq 0$ in aggregate political time series (see, e.g., Box-Steffensmeier and Smith 1996; Lebo, Walker, and Clarke 2000).

[11]To this list, one could add Honaker and King's (2010) imputation in "unbalanced panels," which cannot be used in an RCS design where each case has data missing from every wave but one.

Clarke and Lebo 2003; Lebo, Walker, and Clarke 2000).[12] Fractional integration has also been found in aggregate studies of congressional roll-call data (Lebo, McGlynn, and Koger 2007), Supreme Court decisions (Hurwitz and Lanier 2004; Lanier 2003), crime statistics (Greenberg 2001), and campaign expenditures (Box-Steffensmeier, Darmofal, and Farrell 2009).

The correct level of integration—be it zero, one, or something in between—must be estimated and used to difference a time series in order to get trustworthy inferences. This has been shown from early work by Granger and Newbold (1974) to Box and Jenkins's (1976) techniques to more recent work on fractional integration (Lebo, Walker, and Clarke 2000; Tsay and Chung 2000). To study RCS data at two levels of analysis, the strategy must account for the properties of the aggregate-level (level-2) data.[13] Estimating the (p,d,q) parameters of an ARFIMA model at the aggregate level (level-2) is our first step:[14]

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

where $\overline{Y}_t$ represents the observed mean of all y_i within month 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;[15] ϕ_p represents stationary autoregressive (AR) parameters of order p; θ_q represents q moving average (MA) parameters; and ε_t is a white noise error term for the level-2 disturbances.

The first filter regresses $\bar{Y}_t$ on its noise model (p,d,q) to create $\bar{Y}^*_t$, a stationary series of residuals free of autocorrelation:

$$\bar{Y}^*_t = (1-L)^d \bar{Y}_t \times \frac{(1-\phi_p L^p)}{(1-\theta_q L^q)}. \tag{2}$$

[12]According to Granger's (1980) aggregation theorem, if $\overline{X}_t$ is created by aggregating individual units such that $x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_t$ and $\alpha_i \sim \beta(0,1)$, the heterogeneous autoregressive behavior (note the subscript on α) means that $\overline{X}_t$ will be a fractionally integrated time series. For public opinion, the heterogeneity assumption fits the literature on the distribution of information and political sophistication in the electorate (e.g., Box-Steffensmeier and Smith 1996; Zaller 1992). Heterogeneity in the *types* of voters in each wave of an RCS means that Granger's aggregation theorem (1980) still applies.

[13]A lagged dependent variable does not solve autocorrelation problems when data are fractionally integrated. We find that the lag of a daily mean (e.g., $\bar{Y}^*_{t-1}$) will be insufficient in the RCS case.

[14]A minimum of 50 time-points is a good rule of thumb for our method since estimates of d are less reliable as T drops. We address strategies for shorter T later in the article.

[15]With a larger number of time-points, values for d can be estimated in Stata, RATS, OX, and R. Stata 12 allows the simultaneous estimation of d alongside AR and MA parameters. See the supporting information for R code for our simulations and examples.

Thus, $\bar{Y}_t$ is a function of two components: stochastic $(\bar{Y}^*_t)$ and deterministic $(\bar{Y}'_t)$. So, $\bar{Y}_t - \bar{Y}'_t$ leaves the stochastic component, $\bar{Y}^*_t$, that is, the part of $\bar{Y}_t$ influenced by Xs rather than by its own past history.

For exogenous variables at the aggregate level, the same approach is followed. Where exogenous variables vary within each month,[16] means should be calculated and noise models created for each $\bar{X}_t$. When an exogenous variable varies only across time and not within a time period (e.g., stock prices), one should find the appropriate noise model for it and create Z^*_t, the movement of Z_t not due to the past history of Z.[17] With Z^*_t, $\bar{Y}^*_t$, and $\bar{X}^*_t$, autocorrelation has been modeled on both sides of the equation.

Up to this point, our proposed model simply uses methods shown elsewhere to work for time series: find the appropriate noise model and filter (i.e., Granger and Newbold 1974 for I[0/1]; Box and Jenkins 1976 for $[p,0/1,q]$; Clarke and Lebo 2003 for $[p,d,q]$). But we would also like to study the individual-level data. For that, we marry the logic of fractional differencing with multilevel modeling and move next to the *within*-month study of RCS data. This involves a filter applied to the individual-level data prior to the estimation of an MLM.

Multilevel Models and Modeling Both Dynamic and Static Processes Together

Political scientists have increasingly relied on MLMs to deal with hierarchical data in which "level-1" units are nested within "level-2" structures (Bartels 2009a, 2009b; Gelman et al. 2008). MLMs allow one to analyze how both contextual and unit-specific factors predict a dependent variable (e.g., Gelman and Hill 2007; Steenbergen and Jones 2002). Beyond the substantive motivation, there are also decisive statistical consequences if one ignores a hierarchical structure. Since observations are not independent, the error structure is a problem. For example, if cases are drawn according to geographic areas or regions, the data are no longer conditionally independent and errors are spatially correlated. As a consequence, standard errors will be biased downward and the probability of Type I error increases (Skrondal and Rabe-Hesketh 2004).

[16]We discuss aggregation at the month level to match RCS data like our first example. Of course, our approach is generalizable to other levels of aggregation.

[17]To distinguish the two types of exogenous variables, we use Z for those that vary only over time and X for those that also vary within time-points.

Temporal clustering is, of course, also a problem—error components should be orthogonal to independent variables. This is violated insofar as spatial *or* temporal autocorrelation exists. In an *l*-level model, the residuals should be conditionally independent at *l+1*. This becomes tenuous with time in the model. And, if errors are correlated over time, standard errors will be incorrect.

Various MLMs have led to useful advances with data indexed over time. For one, MLMs have been used to analyze true panel data, where multiple observations are clustered at the country level (Beck 2007; Beck and Katz 2007; Shor et al. 2007).[18] Beck and Katz (2007) note that if one assumes that a dynamic process exists, an LDV can be included in the intercept equation. But where the LDV is not measured, a different solution is needed.

In RCS designs, the individual-level data are nested within multiple, sequential time-points. As with geographic clusters in a single cross-sectional data set, cases are not independently observed. MLMs are well suited for these structures, as individuals can be viewed as embedded within the date the specific cross-section was collected (DiPrete and Grusky 1990).[19] In MLM terms, the individual-level observations, *i*, are the level-1 units and are nested within level-2 units of time, *t*. Once ARFIMA methods are applied at the aggregate level, we must next pay attention to this hierarchical structure.

To fix problems of serial correlation at level-1, we subtract the daily deterministic component from the level-1 dependent variable:

$$y_{it}^{**} = y_{it} - (\bar{Y}_t - \bar{Y}_t^{*}). \tag{3}$$

Note that this step removes the deterministic component from y_{it}, so that y_{it}^{**} now consists of within-month as well as non-temporally autocorrelated between-month variation. We then filter our *x*s through the month-level effects:

$$x_{it}^{**} = x_{it} - \bar{X}_t. \tag{4}$$

The logic is the same as that of Bafumi and Gelman (2006). By accounting for level-1 and level-2 effects, correct parameter estimates can be retrieved.[20]

A multilevel model now puts the double-filtered data to work.[21] The level-1 equation—the within-month model—can be written as

$$y_{it}^{**} = \alpha_{1t} + \beta_1 x_{it}^{**} + u_{1it}. \tag{5}$$

The intercepts, α_{1t}, vary across months where $\alpha_1 \sim N(\alpha_2, \sigma^2)$. In other words, the intercept α_{1t} represents the month-averaged score of y_{it} purged of autocorrelation. It is simply $\bar{Y}_t^{*}$. We can subsequently define these intercepts to be a stochastic function of aggregated individual-level effects, $\bar{X}_t$, and aggregate-level covariates, Z_t:

$$\alpha_{1t} = \alpha_2 + \beta_2 \bar{X}_t^{*} + \gamma Z_t^{*} + u_{2t}. \tag{6}$$

The error terms for Equations (5) and (6) are represented as u_{2t} and u_{1it} for level-1 and level-2 units, respectively.

Combining the equations yields

$$y_{it}^{**} = \alpha_{1t} + \beta_1 x_{it}^{**} + u_{1it} + \beta_2 \bar{X}_t^{*} + \gamma Z_t^{*} + u_{2t}. \tag{7}$$

Where 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 months.

Our model is also well suited for the estimation of time-varying parameters. As we show in our examples below, one can specify coefficients that will vary across time for certain independent variables, w_{it}^{**}. If a level-1 relationship might change in different contexts, a time-varying coefficient, δ_t, can be specified. Thus, Equation (5) can be expanded to

$$y_{it}^{**} = \alpha_{1t} + \beta_1 x_{it}^{**} + \delta_t w_{it}^{**} + u_1. \tag{8}$$

The steps can be summarized as follows: First, create monthly means for the level-1 data of interest, $\bar{Y}_t$ and $\bar{X}_t$. Second, find the proper noise models for them as well as for level-2 series, Z_t, that do not vary within months. Third, filter each through its noise model to create level-2 series free of autocorrelation, $\bar{Y}_t^{*}$, $\bar{X}_t^{*}$, and Z_t^{*}. Fourth, remove month-level deterministic components from the level-1 data. Fifth, estimate an MLM in two levels using the double-filtered data.

Our model offers several advantages. We use the most reliable techniques available—ARFIMA models—to filter out level-2 autocorrelation. Addtionally, by taking the deviations of *i* from level-2 values, we fix problems of serial correlation at level-1. In addition, we are able to include level-2 variables that do not vary within time-points as covariates as well as investigate interesting time-varying effects. In the next section, we use Monte Carlo

[18]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.....n* observed repeatedly over time, *t*. In this case, the country is the level-2 unit, observed repeatedly over time (level-1).

[19]DiPrete and Grusky (1990) propose an MLM for RCS data, but our method is quite different by more fully considering degrees of integration.

[20]To obtain within-month deviations, we remove the random and nonrandom variation in $\bar{X}_t$, where $\overline{X}_t = \bar{X}_t^{*} + \bar{X}_t'$. Thus, $x_{it}^{**} = x_{it} - (\bar{X}_t - \bar{X}_t^{*}) - \bar{X}_t^{*} = x_{it} - \bar{X}_t$.

[21]Our approach is distinct from the differences-in-differences (DID) model frequently used to analyze RCS data, where cross-sections are included before and after a policy intervention (see Athey and Imbens 2006 for a thorough review; see also Wooldridge 2001).

analyses to compare the statistical consequences of our approach to several alternatives.

Simulations

We expect that if the dynamic component is ignored, estimates will be adversely affected, more so as time dependence at level-2 grows. Even if a lag at level-2 is modeled, bias will still be present to the extent that the lag insufficiently accounts for autocorrelated errors. Moreover, if errors are clustered—and thereby not independent—the standard errors will be incorrect.

We simulate data meant to mimic the properties of RCS data.[22] We generated 11,000 data sets, with each consisting of 300 waves and a sample size of 100 per wave.[23] Aggregate values of the independent variable, $\bar{X}_t^*$, were created along with x_{it}^{**} values for within-month variation.[24] Level-1 observations, y_{it}, were generated as a function of individual-level effects, x_{it}^{**} (specified to have a slope coefficient of 0.5), aggregate-level effects, $\bar{X}_t^*$ (specified to have a slope coefficient of 0.3), and random error. 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.[25]

We tested the statistical properties of eight estimation strategies for each data set. We start by presenting "naïve models" where we fail to separate out the aggregate- and individual-level effects. We do this using (1) OLS (labeled here OLS-Naïve) as well as (2) a multilevel model (MLM-Naïve) where intercepts vary across months. We next report the results of six additional strategies that could be used: (3) OLS pooling all data but separating aggregate- and individual-level effects (OLS), (4) OLS specifying aggregate- and individual-level effects and including a month-level lagged dependent variable (OLS-LDV), and (5) OLS accounting for nonstationarity by fractionally differencing the aggregate-level monthly means (ARFIMA-OLS). We also estimated three additional types of MLMs: (6) an MLM separating aggregate- and individual-level effects and allowing intercepts to vary across time (MLM), (7) a version of (6) that adds a month-level lag (MLM-LDV), and (8) fractionally differencing the aggregate series and allowing intercepts to vary across time (ARFIMA-MLM).[26]

Simulation Results

Naïve Models. The naïve models are clearly insufficient.[27] For OLS, the estimates fall between the true slopes with low levels of *d*. But as *d* rises, so does the spread of the estimates, and the average estimated coefficient is biased toward zero. In other words, as the dependent variable becomes less stationary, slopes become more biased and less efficient.

The MLM-Naïve method properly retrieves the individual-level effect, but two problems are present: (1) it ineffectively models aggregate-level processes, since they are inseparable from individual-level effects (see also Bafumi and Gelman 2006; Bartels 2009a); and (2) standard errors are incorrect. As *d* increases, the standard errors are biased downward, leading to incorrect inferences (see Table S3 in the supporting information).

Additional Models. Beyond the naïve models, we need to confront the problems of modeling both individual- and aggregate-level effects together as well as address the likelihood of nonstationarity at level-2. The latter problem has been especially ignored by social scientists.[28] Next, we explore the empirical consequences of six additional approaches: OLS, OLS-LDV, ARFIMA-OLS, MLM, MLM-LDV, and ARFIMA-MLM.

What should we expect from each OLS approach? By pooling all observations and running an OLS regression

[22]Recall that the many recent studies that have carefully analyzed the properties of aggregate-level RCS public opinion data have all found that *d* falls between 0 and 1 and is usually closer to 1 (see Gil-Alana 2008 for a review of this literature). No study we know of has tested for fractional integration in long-T data and found the dichotomous ARIMA approach to be preferable.

[23]We drop the first five waves since there is insufficient data to accurately fractionally difference the first few observations.

[24]Month-level means of *x* are drawn from a standard normal distribution. We then duplicate these observations 100 times to generate a data set size of 30,000. These observations serve as the month-level random noise ($\bar{X}_t^*$). Next, we take a random draw from a normal distribution to generate the within-month independent variable, x_{it}^{**}.

[25]We first calculated the month-level means for y and subtract y_{it} from these values. This gives us the deviations from the month-level mean. Then we fractionally integrated the month-level means and added back in the deviations, giving y_{it}. 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. We use the Hurst R/S statistic (Hurst 1951). The value of *d* equals the Hurst R/S coefficient minus 0.5.

[26]All simulations and analyses were carried out in *R* using restricted maximum likelihood (REML). The MLM models were estimated using the lmer() function in the "lme4" package (Bates and Maechler 2010). Our code is in the supporting information.

[27]See Figure S1 and the discussion that follows it in the supporting information for further explanation.

[28]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.

(solution 3: OLS), we should retrieve incorrect parameter estimates for the aggregate effects of x ($\bar{X}_t$). Simply specifying an OLS model with a lagged aggregate dependent variable, $\bar{Y}_{t-1}$ (OLS-LDV), should produce unbiased and efficient estimates only if the lag accounts for aggregate-level autocorrelation (Achen 2000). Since this will not occur in the presence of fractional integration, bias is expected. The approach of using an ARFIMA model for month-level x ($\bar{X}_t$) and month-level y ($\bar{Y}_t$) and employing OLS will result in incorrect standard errors since OLS cannot effectively account for unobserved aggregate-level variation.

The MLM approaches should be an improvement, but the MLM assumption of independent level-2 errors will be violated insofar as level-2 units are correlated. Thus, an MLM without filters will produce biased and inefficient estimates as d increases. Similarly, an MLM with a lagged dependent variable, $\bar{Y}_{t-1}$, will produce biased estimates and standard errors that are increasingly incorrect as d increases. We expect that the ARFIMA-MLM model will prove to be the most reliable approach.

Figures 1 and 2 display our estimates of bias and inefficiency for the various OLS and MLM models, respectively. Bias is the average of each parameter estimate divided by the true parameter value for each level of d. Thus, a value of 100 indicates a lack of bias. We calculate efficiency as $\sqrt{\frac{\sum(\theta-\bar{\theta})^2}{n}}$, the degree of variation around the average estimate, where n is the number of data sets (1,000). Smaller values indicate greater efficiency. We display the results for four sets of estimates: bias and RMSE for each of the aggregate-level effects (β for $\bar{X}_t$ for the OLS, OLS-LDV, MLM, and MLM-LDV models and β for $\bar{X}_t^*$ for the ARFIMA-OLS and ARFIMA-MLM models) and for the individual-level effects (β for x_{it} for the OLS, OLS-LDV, MLM, and MLM-LDV models and β for x_{it}^{**} for the ARFIMA-OLS and ARFIMA-MLM models).

Figure 1 demonstrates that OLS and ARFIMA-OLS perform reasonably well in terms of retrieving the correct slope for the between-month effect of x on y. ARFIMA-OLS is the most accurate, which is to be expected since it effectively controls for nonstationary month-level effects. OLS-LDV, however, is problematic; the estimated slopes are biased downward as d increases. The upper-right quadrant similarly shows that ARFIMA-OLS has no problems of inefficiency, whereas OLS and OLS-LDV grow more inefficient as d increases.[29] All three methods perform well in terms of retrieving correct individual-level effects, evident in the fact that the lines can barely be discerned in the bottom-left quadrant of Figure 1. So long as one subtracts the month-level means from the observed data, correct within-month parameter estimates can be retrieved. However, the efficiency of estimates is compromised in the case of the OLS models.

Figure 2 presents the diagnostics for the MLM approaches. ARFIMA-MLM estimates of β for $\bar{X}_t^*$ prove to be the best in terms of being unbiased and efficient. In all, the results demonstrate the importance of accounting for nonstationarity.

As one final check on the models, we measure the variation in the standard errors for these models at various levels of d. To this end, we present "optimism" in Table 1, which contrasts the estimated standard errors to sampling variation (see Shor et al. 2007). Following Beck and Katz (1995), $Optimism = 100 \times \sqrt{\frac{\sum_{l=1}^{1000}(\theta_l-\bar{\theta})^2}{\sum_{l=1}^{1000} SE\theta_l}}$. Values greater than 100 indicate that true sampling variation is greater than estimated variation and that 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 above 100 increase Type 1 error rates, the critical inferential problem here.

As Table 1 illustrates, the standard errors are much too small for *all* methods except the ARFIMA-MLM model.[30] At all levels of d, the standard errors are severely "overconfident" in the OLS models. That is, true sampling variability is much larger than estimated variability, leading to t-statistics that are too large. For the OLS and OLS-LDV estimates, this effect grows as d increases. This is also evident for the MLM and MLM-LDV models. As the aggregate means are increasingly a function of past values, the standard errors are underestimated.

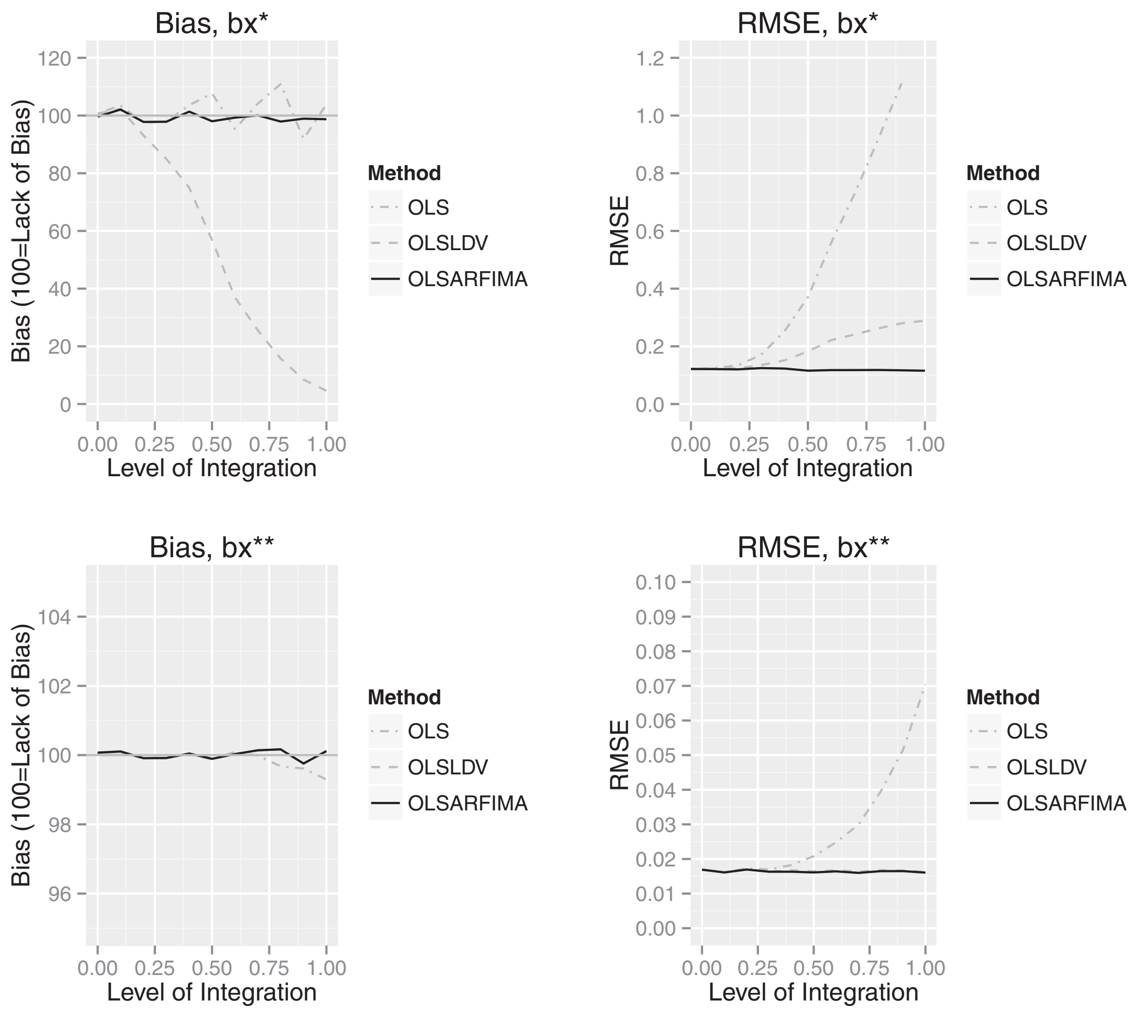
The ARFIMA-OLS models do well at various values of d but have optimism scores that are consistently high due to the model's inattention to aggregate-level variation. Only after accounting for aggregate tendencies as well as individual-level heterogeneity can one retrieve standard errors that reflect true sampling variability. The winner is the ARFIMA-MLM model.

In all, the simulations strongly support the need to consider and model the memory across time-period clusters. With (OLS-LDV and MLM-LDV) or without (OLS and MLM) a lagged cluster mean, the models perform poorly when long memory is ignored. As d increases,

[29]See Figures S2 and S3 in the supporting information for the standard error distributions of each approach under different assumptions about d. The ARFIMA-MLM approach clearly stands out as best.

[30]We present only the optimism estimates for the aggregate-level effects. The optimism estimates hover near 100 for the individual-level effects in all six models.

FIGURE 1 Bias and RMSE for OLS Coefficients



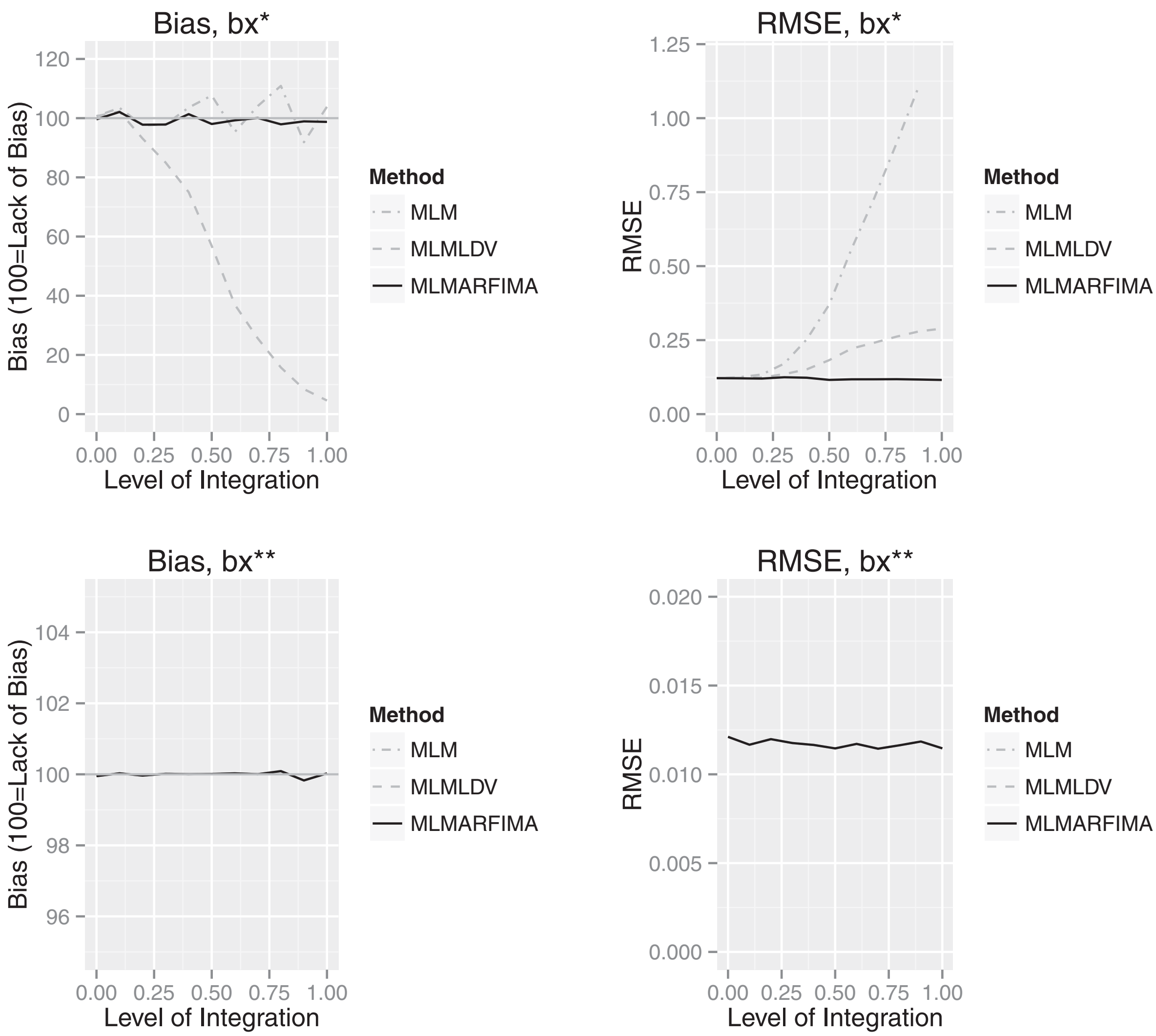
Note: For the OLS and OLS-LDV models, this is the coefficient for $\bar{X}_t$. For the ARFIMA-OLS models, it is the coefficient for $\bar{X}^*_t$. Lines in the bottom panels are all present but overlap.

these models produce biased and less precise parameter estimates and standard errors that are too small. On the other hand, ARFIMA-MLM and ARFIMA-OLS produce unbiased parameter estimates when fractional integration is modeled. Yet, ARFIMA-OLS estimates of aggregate-level standard errors will be too small, elevating the risk of Type I error. Thus, we advocate the ARFIMA-MLM model when cross-sections are related over time.

What to Do When T Is Too Short for ARFIMA

As mentioned, having $T > 50$ is a good rule of thumb for reliable estimates of d and the use of ARFIMA methods. But even with much shorter data sets, an MLM approach can prove useful in overcoming worries of autocorrelation, for examining multiple levels of analysis at once, and for studying time-varying relationships. Shorter RCS examples include the cumulative American National Election Study (see, e.g., Stoker and Jennings 2008), where an MLM might test the effects of context on electoral choice over 20 or so elections. One can apply stationarity tests to such data and generalize our ARFIMA approach to Box and Jenkins's (1976) original ARIMA framework. That is, simpler models are available when d is an integer—or when too few waves exist to properly estimate d as a real number. When diagnosed as stationary ($d = 0$), a series can be modeled as $\bar{Y}_t = \frac{(1-\theta_q L^q)}{(1-\phi_p L^p)}\varepsilon_t$; and, where $d = 1$, Equation (1) simplifies to a differenced version of $\overline{Y}_t$: $\Delta\bar{Y}_t = \frac{(1-\theta_q L^q)}{(1-\phi_p L^p)}\varepsilon_t$. Following that, the second filter can be applied to the individual-level data. If the best model

FIGURE 2 Bias and RMSE for the Random Intercept Multilevel Models



Note: For the MLM and MLM-LDV models, this is the coefficient for $\bar{X}_t$. For the ARFIMA-OLS models, it is the coefficient for $\bar{X}_t^*$. Lines in the bottom panels are all present but overlap.

is simply (0, 0, 0)—that is, no autocorrelation exists in the aggregate—then the model reduces to mean-centering of the level-1 units (as suggested by Bafumi and Gelman 2006).

In cases where T is very short (e.g., less than 10) and tests of stationarity are unreliable, one could simply use theory to decide whether differencing is appropriate at level-2. For example, with 10 monthly waves and a dependent variable of partisanship, assuming long memory and differencing is a better choice than leaving the level-2 data in level form. With longer T, however, it is best to begin with an ARFIMA noise model at level-2. Next, we demonstrate the usefulness of the ARFIMA-MLM approach in three separate examples.

Applications: ARFIMA-MLM at Work

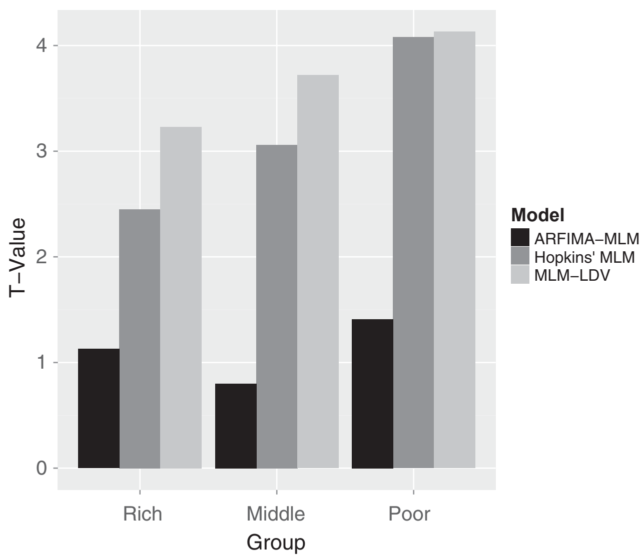
In the three examples that follow, we demonstrate the advantages of our method. First, in a comparison to another MLM approach to RCS data (Hopkins 2012), ours offers improved statistical accuracy and different aggregate-level findings. Second, expanding a strictly time-series approach (Lebo, McGlynn, and Koger 2007), our model allows a richer depth of theoretical development and empirical testing. And third, using the National Annenberg Election Study, our model's flexibility allows the study of a complex campaign environment not fully explored in the literature (e.g., Brady and Johnston 2006; Kenski, Hardy, and Hall Jamieson 2010).

TABLE 1 Optimism Index for Six Modeling Approaches

	Between-Day Effects ($bx*$)					
d	OLS	OLS-LDV	ARFIMA-OLS	MLM	MLM-LDV	ARFIMA-MLM
0	739	1896	737	104	267	104
0.1	783	1865	734	110	262	103
0.2	857	1630	726	118	228	92
0.3	1083	1436	753	145	200	106
0.4	1663	1244	746	214	172	105
0.5	2566	1056	700	312	146	98
0.6	4002	1603	713	456	221	100
0.7	5231	2433	713	565	336	100
0.8	6677	3747	714	698	521	101
0.9	8372	5895	710	856	824	100
1.0	8983	7770	699	907	1088	98

Note: 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 ARFIMA-OLS and ARFIMA-MLM models, they are based on the standard error of the coefficient for $\bar{X}^*_t$.

FIGURE 3 The Effect of Income Growth for 20th Percentile of Income Using Three Modeling Approaches



Note: Each bar represents the *t*-statistics for the respective group's regression coefficient in three modeling approaches (darkest is ARFIMA-MLM, medium is MLM, and lightest is MLM-LDV).

Example 1: What Other MLMs Miss: Hopkins's (2012) "Whose Economy?"

Our first of three examples presents our toughest test. Since others have suggested that RCS data are best dealt with in an MLM format, can we demonstrate ARFIMA-MLM's value over other multilevel approaches? In "Whose Economy? Perceptions of National Economic Performance during Unequal Growth," Hopkins (2012) uses a multilevel model to study over 215,000 respondents nested in 388 months of the Michigan Survey of Consumer Attitudes. These RCS data are a popular source for subjective evaluations of the economy (e.g., Bafumi 2010; Clarke et al. 2005; DeBoef and Kellstedt 2004; Krause 1997; MacKuen, Erikson, and Stimson 1992).

Hopkins's MLM is justified to explain both individual-level factors and their response to aggregate data: "To understand how economic conditions influence Americans' economic perceptions, it is critical to observe attitudes under a range of economic conditions" (2012, 56). Indeed, the key conclusion of the article is at the aggregate level. When asking, "Whose economy matters?" for the formation of economic assessments, "the answer from the SCA is that Americans at all income levels weigh income growth at the low end in their responses" (2012, 68). This is based on regressing sociotropic assessments on aggregate income growth for five groups defined by income percentile.

We replicate Hopkins's analyses and also try the MLM-LDV and ARFIMA-MLM approaches. As Hopkins does, we do this separately for rich, middle-income, and poor respondents. Figure 3 displays the *t*-test statistics for the key finding—the effect of 20th percentile income growth—in three models.[31] Using ARFIMA-MLM, the key effect essentially disappears; light gray bars show the results for Hopkins's method, and the black bars are for ARFIMA-MLM (e.g., $t = 1.41$ for poor respondents).[32]

What the plain MLM misses is that the dependent variable, aggregate sociotropic evaluations, is fractionally integrated (e.g., the *d* estimate is 0.81 for poor respondents). Ignoring between-month dynamics leads to incorrect parameter estimates and an increased likelihood of Type I error.[33] The *t*-statistics associated with income growth at the 20th percentile are considerably greater when ARFIMA filtering is not used. As such, our method

[31]Our code and the full results of all nine models (Tables S5, S6, and S7) are in the supporting information.

[32]For consistency across models, we allow for random effects across months and years. Thus, our estimates are not identical to those reported in Hopkins (2012), but are substantively equivalent. Also, we drop the first five cross-sections from the data (as described in the simulation).

[33]Hopkins's MLM employs a random intercept across years and months, but this does not solve the serial correlation problem. We used this specification, since it is more similar to the MLM-LDV and ARFIMA-MLM models.

gives an inconclusive result as to whether lower-strata income growth shapes national economic evaluations.[34] Recall that the simulations found an optimism index much larger than 100 for $d = 0.8$ with a lagged value of $\bar{Y}$ included. Thus, even with the wealth of data in Hopkins's analyses, neither the MLM nor the MLM-LDV can deal with the data's inherent autocorrelation. The ARFIMA-MLM model (optimism = 100) in such cases is up to the task. The best methods available to time-series analysts need to be married to the MLM approach.

Example 2: Expanding a Time-Series Analysis of Party Unity to an ARFIMA-MLM Setup

In our second example, we show the gains of expanding a strictly macrolevel analysis to include individual-level data: we can specify time-varying effects for covariates and compare effects occurring at two levels. Lebo, McGlynn, and Koger (2007) use ARFIMA techniques to explain how yearly levels of Democratic and Republican Party unity interact closely with each other in congressional roll-call votes from 1789 to 2000. In the Strategic Party Government model, the two parties balance their voting cohesion over time—when a party votes too cohesively, it risks pulling members away from the wishes of their constituents, but a lack of cohesion risks losing important votes and thus electoral support. The strong findings in the aggregate demonstrate this long-term pattern over the course of American history but leave a great deal unanswered. For one, is this strategic behavior a consistently useful way of understanding party behavior on roll calls *within* congresses? Also, has the era of polarization affected the relationship between the parties in their roll-call battles?

The original data are RCS—observations are nested within years but do not repeat over multiple time-points—and we investigate relationships using data at both the roll-call and yearly levels. Beginning with the 29,734 final passage party-line votes from the U.S. House of Representatives nested in 222 years, we calculated year-level means for majority and minority party unity and estimated ARFIMA models for each in order to generate white noise series (i.e., $\bar{X}^*_t$ and $\bar{Y}^*_t$).[35] The second

Table 2 An ARFIMA-MLM Model of Majority Party Voting Unity in the House of Representatives, 1789–2006 (Party Votes Only)

	β (SE)	*t*
Level-1 (within years) N = 29,599		
Intercept	1.390 (0.66)	2.11
Minority unity	0.128 (0.01)	23.15
Policy vote	–6.988 (0.32)	–21.86
Long session	–1.237 (1.37)	–0.90
First session	4.715 (1.33)	3.56
Level-2 (between years) T = 217		
Minority unity	0.557 (0.07)	7.67
Majority cohesion–NOMINATE 1	–44.461 (28.22)	–1.58
Majority cohesion–NOMINATE 2	–42.09 (16.20)	–2.60
Ideological distance–NOMINATE 1	–12.05 (14.68)	–0.82
Ideological distance–NOMINATE 2	1.936 (5.76)	0.34
Majority size	–0.504 (0.10)	–5.28
Error Correction Mechanism (t–1)	–0.226 (0.07)	–3.61
Number of votes/years	29,599/217	

filter subtracts the year-level means from individual observations of x_i and y_i, respectively minority and majority unity on roll-call *i*. With our double filtering done, we add five more variables to the aggregate model (i.e., $\bar{Z}^*_t$): the percent of the House held by the majority and both median-to-median distance between the parties and standard deviations for the majority party in each of the first and second dimensions of DW-NOMINATE scores (Poole and Rosenthal 1997).

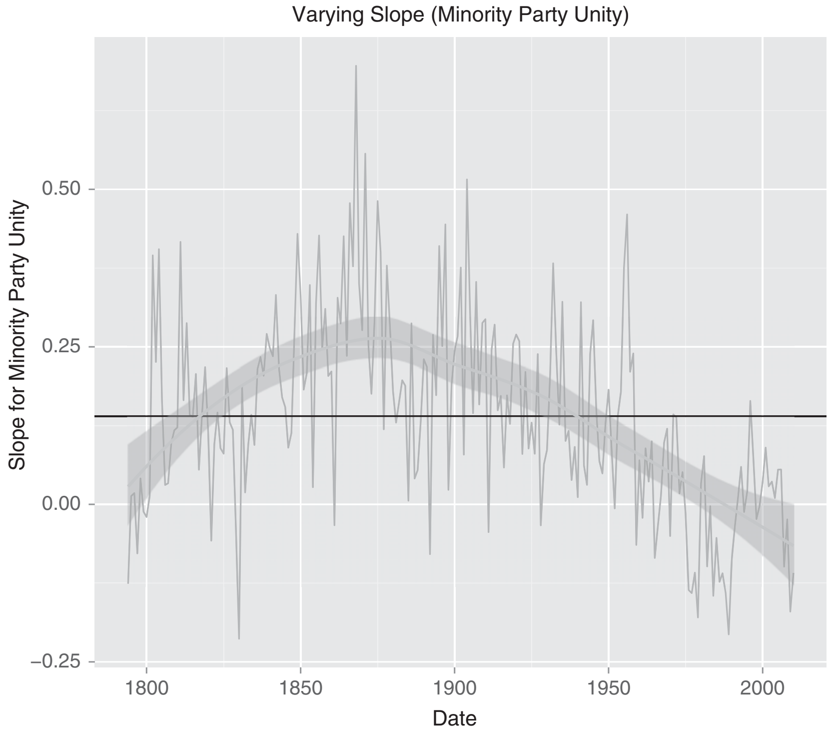
The findings of our ARFIMA-MLM model (Table 2) tell a much more nuanced story than the original one. At level-2, the original relationships all stand up: Majority Party Unity is closely related to Minority Party Unity, close enough that the error correction mechanism is significant. The size of the majority is also important—a larger majority party can afford to be less unified (see also Patty 2008).

Looking at the effect of Minority Unity on final passage votes within years, we still see a very strong relationship ($t = 23.15$). Next, we estimate a random coefficient for Minority Party Unity's year-by-year effect at level-1. The solid line in Figure 4 gives us the overall coefficient, whereas the jagged line gives us the yearly value (and Loess smoother) for δ_t, the effect of Minority Party Unity on Majority Party Unity for individual roll-call votes. The relationship obviously varies a great deal over time, and

[34] In the supporting information, we present models restricted to include only the 95th and 20th percentiles of income growth. The substantive results remain identical—the *t*-statistics do not reach conventional levels of significance across income categories.

[35] Both majority and minority party unity are fractionally integrated, with *d* values of 0.77 and 0.81, respectively. See the supporting information for further discussion.

FIGURE 4 The Time-Varying Effect of Minority Unity on Majority Unity, 1794–2006



Note: Coefficient and smoother for the roll-call-level effect of Minority Party Unity on Majority Party Unity.

pooling would hide this entirely.[36] In particular, the strength of the relationship has been dropping for some time and is now essentially zero. In an increasingly polarized political environment, there is extremely little variation in unity across votes.[37] As members of Congress appeal to increasingly polarized constituencies, the strategic adjustments parties make are primarily aggregate shifts from one year to the next, rather than within-Congress variation across roll-call votes.

Unlike the Hopkins (2012) example, here the original study used only aggregate data and accounted for autocorrelation in the data. However, it was limited in scope in that the level of analysis was strictly aggregate. The use of ARFIMA-MLM opens up possibilities for studying the thousands of individual roll calls that compose the complete RCS data set. A similar approach could extend studies of key dependent variables in public opinion, Supreme Court decisions, or conflicts nested in time, for example.[38]

Example 3: Modeling Dynamic Effects during a Campaign

The 2008 presidential election campaign was a complex one. A full and flexible model of vote choice or candidate evaluation should be able to account for the effects of the unfolding economic crisis as well as time-varying individual-level factors such as sociodemographics and economic and political judgments. Kenski, Hardy, and Hall Jamieson (KHHJ 2010) study the election from multiple angles using the daily data of the National Annenberg Election Survey (NAES). In some dynamic analyses, they demonstrate how assessments of the parties, candidates, and issues changed—sometimes dramatically—over the course of the campaign (e.g., 2010, 4, 18, 19, 93, 156). Elsewhere, they pool together thousands of respondents interviewed over several weeks and run static

[36]Tests for time-varying slopes: no random slope (AIC = 276,331, BIC = 276,447, –LL = 138,151), random slope (AIC = 275,441, BIC = 275,574, –LL = 137,705), $p < .01$ for the likelihood ratio test.

[37]See the supporting information for graphs of the data, the distribution of Minority Party Unity and Majority Party Unity by decade, and further details.

[38]See Table S1 in the supporting information for a list of recent examples where our approach could prove useful.

TABLE 3 An ARFIMA-MLM Model of Campaign Effects

	Within Day		Between Day	
	B **(SE)**	**t**	***B*** **(SE)**	**t**
Fundamentals				
Intercept	–0.09 (0.11)	–0.77		
Party identification	0.13 (0.01)	10.09	–0.74 (0.63)	–1.16
Ideology (conservative)	–0.06 (0.02)	–3.28	–1.01 (0.78)	–1.29
Vote Bush in 2004	–0.00 (0.05)	–0.10	—	
Approve Bush	–0.36 (0.05)	–7.03	2.65 (2.44)	1.09
National economy	–0.11 (0.05)	–2.41	0.14 (2.02)	0.07
Personal economy	–0.02 (0.05)	–0.48	–4.67 (2.33)	–2.00
Sociodemographics				
Gender (female)	–0.03 (0.03)	–0.79		
Age (in years)	0.004 (0.001)	2.86		
Black	0.15 (0.07)	2.19		
Hispanic	0.16 (0.17)	0.97		
Education	–0.005 (0.007)	–0.70		
Income (in thousands)	0.0003(0.0004)	0.71		
Media				
Number of days saw campaign info (TV)	0.02 (0.008)	2.31	0.46 (0.30)	1.34
Number of days heard about campaign (radio)	0.02 (0.006)	3.25	0.19 (0.29)	0.68
Number of days saw info (newspaper)	0.008 (0.006)	1.34	0.35 (0.26)	1.35
Number of days saw info (Internet)	0.008 (0.006)	1.34	–0.49 (0.25)	–1.95
Campaign Messages				
Elect McCain is like reelecting Bush	0.65 (0.05)	13.87	0.47 (2.20)	0.21
McCain is too old	0.37 (0.04)	8.73	0.65 (1.72)	0.38
Obama's ideology (liberal)	–0.05 (0.02)	–2.33	0.60 (1.03)	0.59
Experience (McCain-Obama)	–0.14 (0.01)	17.29	0.23 (0.37)	0.63
Judgment (McCain-Obama)	–0.24 (0.01)	–29.90	0.64 (0.40)	1.58
Patriotic (McCain-Obama)	–0.12 (0.01)	–16.92	–0.57 (0.36)	–1.60
Values (McCain-Obama)	–0.36 (0.01)	–47.61	–1.35 (0.38)	–3.52

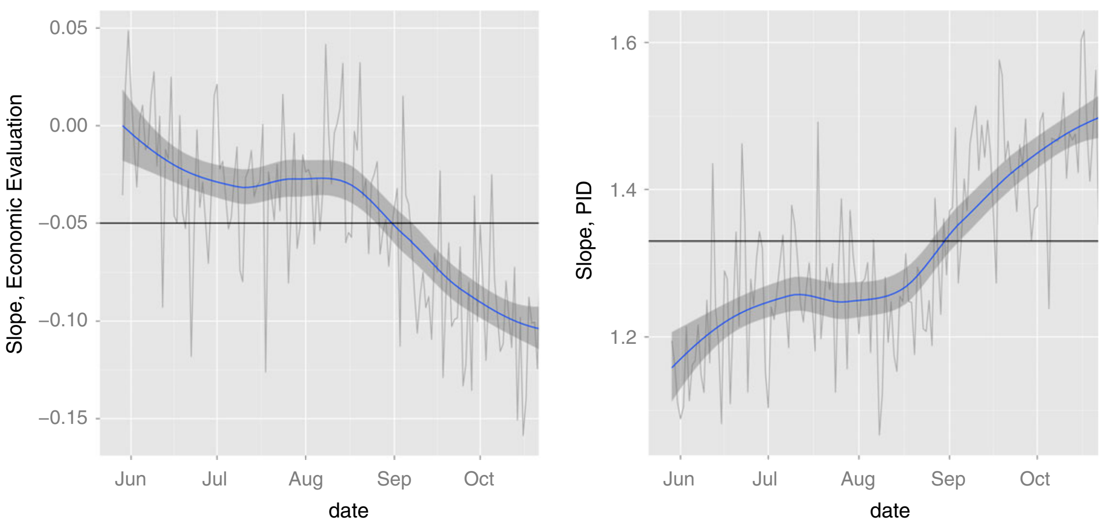
Note: For consistency, we present the random intercept model. The intra-class correlation is small (<0.01), and the OLS-ARFIMA model separating the within- and between-day effects yields equivalent results. The model is similar to Kenski, Hardy and Hall Jamieson's (2010), with several qualifications. Rather than vote choice, the dependent variable is Evaluation of Obama–Evaluation of McCain.
Source: 2008 NAES.

analyses (e.g., 2010, 168, 271, 275, 299, 316). Aside from the statistical problems of pooling, KHHJ neglect to exploit the data's capability to simultaneously study how time-varying individual-level factors and dynamic campaign-level processes influence political behaviors and judgment.

In our final example, we use ARFIMA-MLM to assess the roles played by the "fundamentals" and campaign-specific characteristics while taking account of the (perhaps time-varying) voter-level factors that affected evaluations of Senators Obama and McCain. Our dependent variable measures positive versus negative evaluations of Barack Obama relative to John McCain, with 0 as a midpoint and higher values indicating pro-Obama sentiment ($M = 0.16$, $SD = 4.67$).[39] In the aggregate, the dependent variable is demonstrably nonstationary ($d = 0.90$), thus

[39]We include observations beginning with Senator Obama's capturing of the Democratic nomination in June through to Election Day. Except for four questions that lack observations prior to October 2008 ("believing that Obama will better handle the economy," rating of Sarah Palin as "not ready to be president," rating of Joseph Biden as "ready to be president," and believing that "Obama will raise my taxes but McCain will not"), our covariates are the same at KHHJ's model (2010, 299–300). However, we do not use their dichotomous dependent variable, vote choice. The intuition and ARFIMA filtering are the same for a dichotomous dependent variable, but the method involving our second filter is different.

FIGURE 5 Slopes for Sociotropic Economic Evaluations and Party Identification over the 2008 Campaign



making our ARFIMA filtering especially necessary.[40] On the right-hand side, our covariates follow KHHJ's (2010, 299) specification and include four categories: fundamentals, sociodemographics, media, and opinions on central campaign messages.

Our ARFIMA-MLM model provides estimates for within-day and between-day effects in Table 3. The statistically significant findings of the original model are italicized. Several important characteristics emerge and demonstrate the difference in aggregate- and individual-level inferences. For example, an optimistic assessment of the state of the national economy translates into more positive McCain evaluations ($b = -0.11$, SE = 0.05, $p < .05$), but as the electorate's opinion varied over the campaign, this did not affect candidate evaluation in the aggregate ($b = 0.14$, SE = 2.02, ns). The reverse is true for personal economic evaluations: ratings of respondents' own financial situation have a nonsignificant impact on evaluations ($b = -0.02$, SE = 0.05, ns), whereas the aggregate movement of the variable led to more positive McCain assessments ($b = -4.67$, SE = 2.23, $p < .05$). Certainly, the effects of the economy and economic judgments on political evaluations are complex. Using RCS data to the fullest extent prevents us from oversimplifying the relationships present in the data.

Lastly, in a differently specified model, we examine how economic evaluation and party identification shape candidate evaluation. We estimate one model predicting candidate evaluation in which the slopes are fixed across days, and a second model allowing the within-day slopes, δ_t, to vary across days for economic evaluations and party identification.[41] The time-varying coefficients (plus a Loess smoother) are plotted in Figure 5, with the solid lines representing the constant coefficient we would get if we lose the flexibility of estimating δ_t. In line with the notion that campaigns activate and reinforce latent predispositions (Lazarsfeld, Berelson, and Gaudet 1944), we see these effects grow significantly over the five-month period. In all, this example gives us a nice array of the advantages of the ARFIMA-MLM approach: (1) individual- and aggregate-level relationships are complex and can be studied simultaneously, (2) the time-varying effects of covariates can be studied, and (3) dealing with level-2 nonstationarity makes for trustworthy statistical inferences.

Conclusion

Given what is known about the problems of time-series data and our investigations here, time-level clustering is an important issue to consider in RCS data sets. Most of the prior work using RCS data has been unsatisfactory—analyzing either within or between processes exclusively, not both simultaneously. We demonstrate that failing to account for dynamic effects that exist can lead to biased parameter estimates and incorrect standard errors. Our

[40]We follow KHHJ's practice of calculating five-day rolling averages for the aggregated series to overcome the noisiness of the day-level aggregates caused by the small daily samples.

[41]We refer to these two models as Example 3A and 3B in the supporting information.

solution is a two-step filtering process where means are retrieved, a level-2 ARFIMA model is specified, and then level-1 data are filtered through these estimates. Each of the other seven approaches we test encounters problems in one area or another, but our ARFIMA-MLM always performs well, especially as memory in the aggregates gets longer.

In addition, as demonstrated in our last two examples, we 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 constants 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 dynamic contextual factors. Even so, without a method such as double filtering, the inferences from such an exercise would be suspect.

It is worth noting additional points of flexibility within our framework. Where T is lower, one can switch to models that do not involve estimating the fractional integration parameter, *d*. ARMA and ARIMA models are just particular types of ARFIMA models but are safer to estimate up until the point of about $t = 50$ (Dickinson and Lebo 2007). Estimating a (p, 0/1, q) model and then filtering values at level-2 will allow the same implementation of our method one could get with longer T.

Additionally, our ARFIMA-MLM framework can certainly be extended to PCSTS designs, but with two notable caveats. First, the number of pseudo-waves that can be compiled into an RCS design may be quite high, perhaps running into the hundreds of consecutive waves. With PCSTS, however, long data sets are rarer. Yearly data by country often top 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 a shorter t, the PCSTS analyst is best served by estimating a simpler noise model at the aggregate level, but the rest of our framework would still apply.

Second, the PCSTS analyst has other methods available to control for autocorrelation, such as panel-corrected standard errors (Beck and Katz 1995). Differencing or using a lagged dependent variable are two imperfect solutions, but they are still improvements on simply pooling the data or ignoring the sequence of the waves. Or, by differencing all the variables in a dynamic panel (Baltagi 2005), unit-specific idiosyncrasies can drop out of a PCSTS model. Our double filtering method can be added to this list of solutions but, admittedly, has more competition in that particular toolbox.

We encourage researchers to adopt the ARFIMA-MLM model when analyzing RCS data or long panels and pseudo-panels. By using multilevel models to study data nested in time, 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 crucial to consider, and with greater attention to simultaneously modeling static and dynamic processes, this will provide a richer depiction of 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 American Political Science Association.

Athey, Susan, and Guido Imbens. 2006. "Identification and Inference in Nonlinear Difference in Differences Models." *Econometrica* 74(2): 431–97.

Bafumi, Joseph. 2010. "Animal Spirits: The Effect of Economic Expectations on Economic Output." *Applied Economics* 43(25): 3573–89.

Bafumi, Joseph, and Andrew Gelman. 2006. "Fitting Multilevel Models When Predictors and Group Effects Correlate." Presented at the annual meeting of the Midwest Political Science Association, Chicago, IL.

Baltagi, Badi. 2005. Econometric Analysis of Panel Data, 3rd Edition. West Sussex: John Wiley & Sons Ltd.

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(3): 474–95.

Bates, Douglas, and Martin Maechler. 2010. "lme4: Linear Mixed Effects Models Using S4 Classes." R package version 0.999375–37. http://CRAN.R-project.org/package=lme4.

Beck, Nathaniel. 2007. "From Statistical Nuisances to Serious Modeling: Changing How We Think about the Analysis of Time-Series-Cross-Section Data." *Political Analysis* 15(2): 97–100.

Beck, Nathaniel, and Jonathan Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross- Section Data." *American Political Science Review* 89(3): 634–47.

Beck, Nathaniel, and Jonathan Katz. 2007. "Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments." *Political Analysis* 15(2): 182–95.

Blaydes, Lisa, and Eric Chaney. 2013. "The Feudal Revolution and Europe's Rise: Political Divergence of the Christian West and the Muslim World before 1500 CE." *American Political Science Review* 107(1): 16–34.

Box, George, and Gwilym Jenkins. 1976. *Time Series Analysis, Forecasting, and Control.* San Francisco, CA: Holden Day.

Box-Steffensmeier, Janet M., David Darmofal, and Christian A. Farrell. 2009. "The Aggregate Dynamics of Campaigns." *Journal of Politics* 71(1): 309–23.

Box-Steffensmeier, Janet M., Suzanna DeBoef, and Tse-min Lin. 2004. "The Dynamics of the Partisan Gender Gap." *American Political Science Review* 98(3): 515–28.

Box-Steffensmeier, Janet M., and Renee M. Smith. 1996. "The Dynamics of Aggregate Partisanship." *American Political Science Review* 90(3): 567–80.

Box-Steffensmeier, Janet M., and Renee M. Smith. 1998. "Investigating Political Dynamics Using Fractional Integration Methods." *American Journal of Political Science* 42(2): 661–89.

Brady, Henry, and Richard Johnston. 2006. *Capturing Campaign Effects.* Ann Arbor: University of Michigan Press.

Brown, David S., and Ahmed M. Mobarak. 2009. "The Transforming Power of Democracy: Regime Type and the Distribution of Electricity." *American Political Science Review* 103(2): 193–213.

Byers, David, James Davidson, and David Peel. 2000. "The Dynamics of Aggregate Political Popularity: Evidence from Eight Countries." *Electoral Studies* 19(1): 49–62.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan. 2002. "Out of Step, out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(1): 127–40.

Carey, Sean D., and Matthew J. Lebo. 2006. "Election Cycles and the Economic Voter." *Political Research Quarterly* 59(4): 543–56.

Clarke, Harold, and Matthew Lebo. 2003. "Fractional (Co)integration and Governing Party Support in Britain." *British Journal of Political Science* 33(2): 283–301.

Clarke, Harold, Marianne Stewart, Mike Ault, and Euel Elliott. 2005. "Men, Women and the Political Economy of Presidential Approval." *British Journal of Political Science* 35(1): 31–51.

DeBoef, Suzanna, and Paul M. Kellstedt. 2004. "The Political (and Economic) Origins of Consumer Confidence." *American Journal of Political Science* 48(4): 633–49.

Dickinson, Matthew J., and Matthew J. Lebo. 2007. "Reexamining the Growth of the Institutional Presidency, 1940–2000." *Journal of Politics* 69(1): 206–19.

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(4): 723–41.

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.

Gil-Alana, Luis A. 2008. "Fractional Integration and Structural Breaks at Unknown Periods of Time." *Journal of Time Series Analysis* 29(1): 163–85.

Granger, Clive W. J. 1980. "Long Memory Relationships and the Aggregation of Dynamic Models." *Journal of Econometrics* 14(2): 227–38.

Granger, Clive, and Paul Newbold. 1974. "Spurious Regression in Econometrics." *Journal of Econometrics* 2(1): 111–20.

Green, Donald, and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science.* New Haven: Yale University Press.

Greenberg, David F. 2001. "On Theory, Models, Model Testing, and Estimation." *Journal of Quantitative Criminology* 17(4): 409–22.

Honaker, James, and Gary King. 2010. "What to Do about Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54(2): 561–81.

Hopkins, Daniel J. 2012. "Whose Economy? Perceptions of National Economic Performance during Unequal Growth." *Public Opinion Quarterly* 76(1): 50–71.

Hurst, H. Edwin. 1951. "Long-Term Storage Capacity of Reservoirs." *Transactions of the American Society of Civil Engineers* 116: 770–99.

Hurwitz, Mark S., and Drew N. Lanier. 2004. "I Respectfully Dissent: Consensus, Agendas, and Policy-Making on the U.S. Supreme Court, 1888–1999." *Review of Policy Research* 21(3): 429–45.

Jerit, Jennifer, and Jason Barabas. 2012. "Partisan Perceptual Bias and the Information Environment." *Journal of Politics* 74(3): 672–84.

Jerit, Jennifer, Jason Barabas, and Toby Bolsen. 2006. "Citizens, Knowledge, and the Information Environment." *American Journal of Political Science* 50(1): 266–82.

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(2): 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* 88(3): 923–37.

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.

Kramer, Gerald H. 1983. "The Ecological Fallacy Revisited: Aggregate-Level versus Individual- Level Findings on Economics and Elections, and Sociotropic Voting." *American Political Science Review* 77(1): 92–111.

Krause, George A. 1997. "Voters, Information Heterogeneity, and the Dynamics of Aggregate Economic Expectations." *American Journal of Political Science* 41(4): 1170–1200.

Lanier, Drew N. 2003. *Of Time and Judicial Behavior: Time Series Analysis of United States Supreme Court Agenda-Setting and Decision Making, 1888–1997.* Selinsgrove, PA: Susquehanna University Press.

Lau, Richard, David J. Andersen, and David Redlawsk. 2008. "An Exploration of Correct Voting in Recent U.S. Presidential Elections." *American Journal of Political Science* 52(2): 395–411.

Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. 1944. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Lebo, Matthew J., Adam J. McGlynn, and Greg Koger. 2007. "Strategic Party Government: Party Influence in Congress, 1789–2000." *American Journal of Political Science* 51(3): 464–81.

Lebo, Matthew J., Robert Walker, and Harold Clarke. 2000. "You Must Remember This: Dealing with Long Memory in Political Analyses." *Electoral Studies* 19(1): 31–48.

MacKuen, Michael, Robert S. Erikson, and James Stimson. 1989. "Macropartisanship." *American Political Science Review* 83(4): 729–50.

MacKuen, Michael B., Robert S. Erikson, and James A. Stimson. 1992. "Peasants or Bankers? The American Electorate and the U.S. Economy." *American Political Science Review* 86(3): 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* 58(1): 169–200.

Moy, Patricia, Michael Xenos, and Verena K. Hess. 2006. "Priming Effects of Late-Night Comedy." *International Journal of Public Opinion Research* 18(2): 198–210.

Patty, John W. 2008. "Equilibrium Party Government." *American Journal of Political Science* 52(3): 636–55.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political Economic History of Roll Call Voting*. New York: Oxford University Press.

Romer, David. 2006. *Capturing Campaign Dynamics, 2000 and 2004: The National Annenberg Election Survey*. Philadelphia: University of Pennsylvania Press.

Segal, Jeffrey A., and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. New York: Cambridge University Press.

Shor, Boris, Joseph Bafumi, Luke Keele, and David Park. 2007. "A Bayesian Multilevel Modeling Approach to Time Series Cross-Sectional Data." *Political Analysis* 15(2): 165–81.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

Steenbergen, Marco R., and Bradford S. Jones. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46(1): 218–37.

Stoker, Laura, and M. Kent Jennings. 2008. "Of Time and the Development of Partisan Polarization." *American Journal of Political Science* 52(3): 619–35.

Stroud, Natalie. 2008. "Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure." *Political Behavior* 30(3): 341–66.

Tsay, Wen-Jen, and Ching-Fan Chung. 2000. "The Spurious Regression of Fractionally Integrated Processes." *Journal of Econometrics* 96(1): 155–82.

Voeten, Erik. 2008. "The Impartiality of International Judges: Evidence from the European Court of Human Rights." *American Political Science Review* 102(4): 417–32.

Wood, B. Dan. 2009. *The Myth of Presidential Representation*. New York: Cambridge University Press.

Wooldridge, Jeffrey. 2001. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1: Coefficient Estimates for Naïve Pooling
Figure S2: Standard Errors for βx^*, OLS Models*
Figure S3: Standard Errors for βx^*, Multilevel Models*
Figure S4: T-statistics for 20th percentile income growth among low, middle, and high income respondents. Estimates are from Table S8.
Figure S5: The Level of Party Unity in the House of Representatives, 1789-2006
Figure S6: The Yearly Standard Deviation of Party Unity in the House of Representatives, 1789-2006
Table S1: APSR and AJPS Papers 2010-2013 where ARFIMA-MLM could be applied
Table S2: Bias of Six Different Estimation Approaches
Table S3: RMSE of Six Different Estimation Approaches
Table S4: Optimism of Six Different Estimation Approaches
Table S5: Three MLM Approaches to Study Retrospective Sociotropic Assessments: Low Income Respondents.
Table S6: Three MLM Approaches to Study Retrospective Sociotropic Assessments: Middle Income Respondents.
Table S7: Three MLM Approaches to Study Retrospective Sociotropic Assessments: High Income Respondents.
Table S8: ARFIMA analysis; only including growth at the 95th and 20th percentiles. Intercepts vary across both months and years. For the ARFIMA-MLM we apply ARFIMA filtering to the DV, national unemployment, gas prices, and oil prices, all of which were then subsequently lagged. Since income growth and stock prices are already first differenced, we leave these in the form reported in Hopkins (2012). Finally, we separate out changes in income levels by creating a month level aggregate of individual income (i.e., aggregate income), which also is fractionally differenced in the ARFIMA-MLM.