A Solution to the Repeated Cross Sectional Design

Matthew Lebo
Stony Brook University
and
Christopher Weber
Louisiana State University

July 28, 2011
28th Annual Summer Meeting of the Society for Political Methodology

Panels, Pseudo-panels, and RCS Designs

- Panels have the same observations at multiple points in time.
- Pseudo-panels do not have identical sets of cases at every point in time.
  - unbalanced panels will have some observations appearing more than once.
  - repeated cross-sectional designs (RCS) will not have any observation appearing more than once.

How prevalent are RCS data?

- Very. For example:
- Cumulative NES file.
- National Annenberg Election Study.
- General Social Survey.
- Stringing together archived files at ICPSR or Roper can create hundreds of consecutive Gallup Surveys, CBS/NYT polls, World Value Surveys.
- Michigan's Survey of Consumers.
- 2004-2009, 68 articles in the APSR and AJPS that use RCS data at individual-level.

A True Panel

t=1	t=2	t=3	…	t=T
$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	…	$y_{1,T}$
$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	…	$y_{2,T}$
$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	…	$y_{3,T}$
$y_{4,1}$	$y_{4,2}$	$y_{4,3}$	…	$y_{4,T}$
…	…	…	…	…
$y_{n,1}$	$y_{n,2}$	$y_{n,3}$	…	$y_{n,T}$

The same person/ country/ state

A Repeated Cross Section Design

Individuals nested in time.

t=1	t=2	t=3	…	t=T
$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	…	$y_{1,T}$
$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	…	$y_{2,T}$
$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	…	$y_{3,T}$
$y_{4,1}$	$y_{4,2}$	$y_{4,3}$	…	$y_{4,T}$
…	…	…	…	…
$y_{n,1}$	$y_{n,2}$	$y_{n,3}$	…	$y_{n,T}$

Not the same person

$y_{1,1}$ indicates person 1 in wave 1 which occurs at t=1.

Option 1: Go Aggregate

t=1	t=2	t=3	…	t=T
$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	…	$y_{1,T}$
$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	…	$y_{2,T}$
$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	…	$y_{3,T}$
$y_{4,1}$	$y_{4,2}$	$y_{4,3}$	…	$y_{4,T}$
…	…	…	…	…
$y_{n,1}$	$y_{n,2}$	$y_{n,3}$	…	$y_{n,T}$
$\overline{Y}_1$	$\overline{Y}_2$	$\overline{Y}_3$	…	$\overline{Y}_T$

Use daily means for time series analysis

Reduces a sample of size N*T to one simply T long.

Traditional "long-t" Time Series

- Model $\overline{Y}_t$ using $\overline{X}_t$.
- Allows study of political dynamics.
- Overcome partisan bias at individual-level (Kramer 1983) and use averages in which bias cancels out.
- Allow use of time series methods like cointegration and time-varying parameters.
- Many key pieces to understanding public opinion over time.

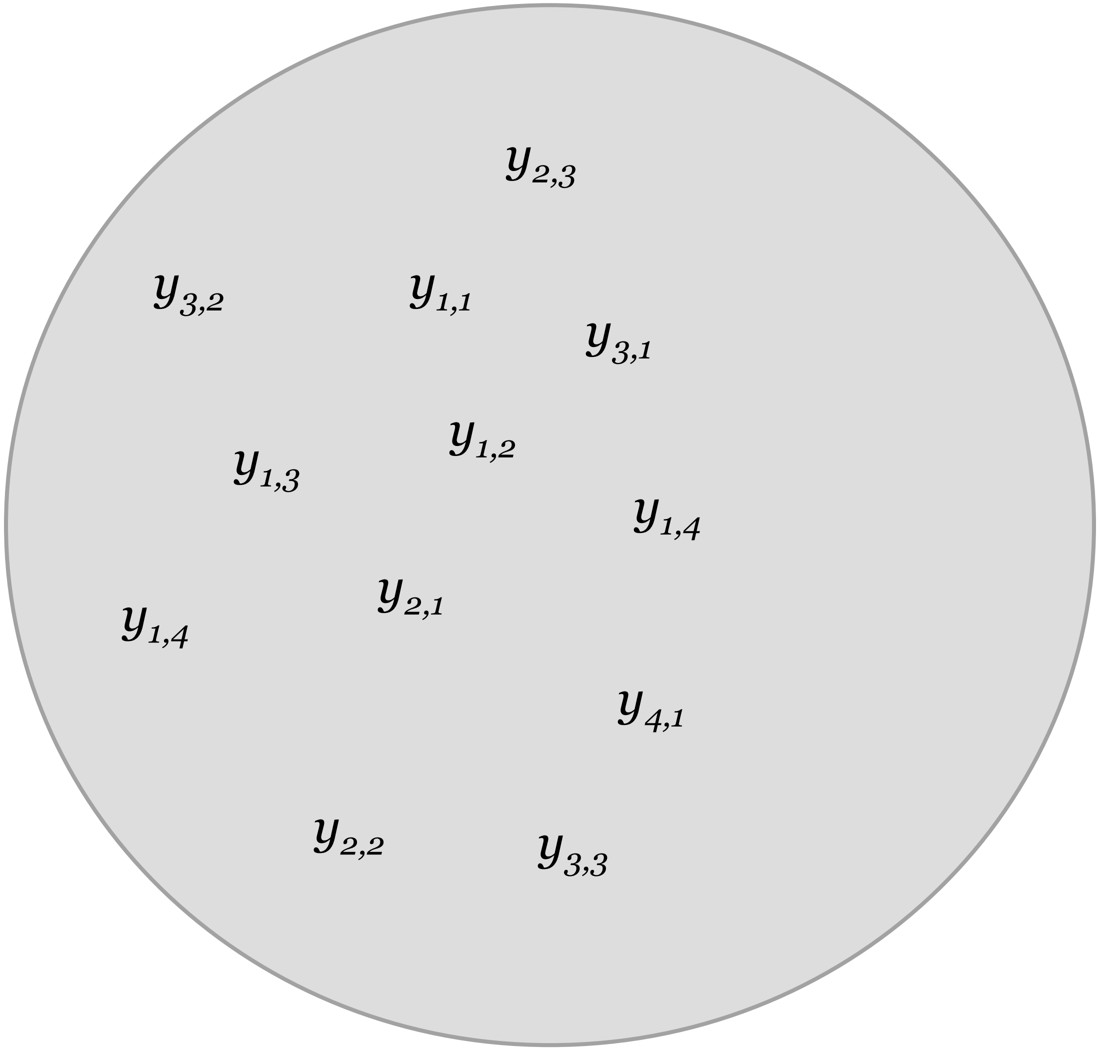
- Examples that begin with RCS and create time series:
  - Mackuen, Erikson, and Stimson (1989; 1992). Gallup Polls.
  - Box-Steffensmeier, DeBoef and Lin (2004). CBS/NYT Polls.
  - Clarke, Stewart, Ault and Elliott (2005). Michigan's Survey of Consumers.
  - Johnston, Hagen, and Jamieson (2004). NAES.
  - Clarke and Lebo (2003). British Gallup.

Another Option: Naïve Pooling

Throw all the cases in together and ignore the time component.

E.g. Romer (2006); Moy, Xenos, and Hess (2006); Stroud (2008).

Confined to cross-sectional hypotheses – no dynamics.



Autocorrelation in a True Panel

t=1	t=2	t=3	…	t=T
$\varepsilon_{1,1}$	$\varepsilon_{1,2}$	$\varepsilon_{1,3}$	…	$\varepsilon_{1,T}$
$\varepsilon_{2,1}$	$\varepsilon_{2,2}$	$\varepsilon_{2,3}$	…	$\varepsilon_{2,T}$
$\varepsilon_{3,1}$	$\varepsilon_{3,2}$	$\varepsilon_{3,3}$	…	$\varepsilon_{3,T}$
$\varepsilon_{4,1}$	$\varepsilon_{4,2}$	$\varepsilon_{4,3}$	…	$\varepsilon_{4,T}$
…	…	…	…	…
$\varepsilon_{n,1}$	$\varepsilon_{n,2}$	$\varepsilon_{n,3}$	…	$\varepsilon_{n,T}$

Correlated due to factors specific to time-point *t*
corr($\varepsilon_{i,t}$, $\varepsilon_{j,t}$)≠0
Thus, solutions like fixed effects and PCSE.

Correlated due to factors specific to individual *i*
corr($\varepsilon_{i,t}$, $\varepsilon_{i,t+1}$)≠0
Thus, solutions like lagged dependent variables and differencing.

What changes with RCS?

- Importantly, the range of solutions.
  - can't use a lagged dependent variable since $y_{i,t-1}$ doesn't appear in the data set.
  - can't difference the dependent variable for the same reason.
  - And, even if you could do either of the above, the methods may be insufficient to account for between wave memory.
  - Panel Corrected Standard Errors premised on a true panel and doesn't solve bias if it exists.
  - Fixed effects don't solve autocorrelation in either direction.
- Does the problem of autocorrelation go away since each observation appears only once?
  - NO!

Autocorrelation in a Repeated Cross Section Design

t=1	t=2	t=3	…	t=T
$\varepsilon_{1,1}$	$\varepsilon_{1,2}$	$\varepsilon_{1,3}$	…	$\varepsilon_{1,T}$
$\varepsilon_{2,1}$	$\varepsilon_{2,2}$	$\varepsilon_{2,3}$	…	$\varepsilon_{2,T}$
$\varepsilon_{3,1}$	$\varepsilon_{3,2}$	$\varepsilon_{3,3}$	…	$\varepsilon_{3,T}$
$\varepsilon_{4,1}$	$\varepsilon_{4,2}$	$\varepsilon_{4,3}$	…	$\varepsilon_{4,T}$
…	…	…	…	…
$\varepsilon_{n,1}$	$\varepsilon_{n,2}$	$\varepsilon_{n,3}$	…	$\varepsilon_{n,T}$

Still correlated due to factors specific to time-point *t*

But, is $\varepsilon_{1,1}$ still too correlated with $\varepsilon_{1,2}$ even when they aren't the same individuals?

Absolutely!

Rephrase that last question

- Do we expect correlations between $\bar{Y}_t$ and $\bar{Y}_{t+1}$?
- Sure. Dozens of papers establish the need to account for memory in time series measured at the aggregate level.
  - Box-Jenkins techniques introduced in the 1970s suggest (*p*,*q*) ARMA models.
  - Clarke and Stewart (1994) introduce unit roots and (*p*,0/1,*q*) ARIMA models.
  - Box-Steffensmeier and Smith (1996) introduce fractional integration and (*p*,*d*,*q*) ARFIMA models.
  - ARFIMA models found to be necessary for many studies since.

Rephrase that last question II

ARFIMA models found to be necessary for many studies.

- Lebo, Walker, and Clarke (2000): Public Mood, presidential approval, Macropartisanship.
- Box-Steffensmeier and Tomlinson (2000): ICS and congressional approval.
- Byers, Davidson, and Peele (2000): approval and party support in many European democracies.
- Box-Steffensmeier, DeBoef and Lin (2004): the gender gap.
- Clarke and Lebo (2003): British party variables, vote intentions, PM approval.
- Box-Steffensmeier and DeBoef (2003): Micro-ideology.
- Treisman (2011) party support in Russia.

And…

- If $\bar{Y}_t$ and $\bar{Y}_{t+1}$ are correlated, then $\varepsilon_{i,t}$ is still correlated with $\varepsilon_{j,t+1}$ more than $\varepsilon_{i,t}$ is correlated with $\varepsilon_{k,t+2}$.
- This is true since observations in each time point are dispersed around a mean correlated with the mean of the adjacent time-point.
- Put another way: $\bar{Y}_t = E(y_{i,t})$ and $\bar{Y}_{t+1} = E(y_{i,t+1})$.

  If corr($\bar{Y}_t$ and $\bar{Y}_{t+1}$)≠0

  then corr($E(y_{i,t})$, $E(y_{i,t+1})$) ≠0 and

  corr($E(\varepsilon_{i,t})$, $E(\varepsilon_{i,t+1})$) ≠0.

How do we deal with these two types of autocorrelation?

- Available methods don’t provide a solution.
  - Can’t difference, can’t use lagged dependent variable.
  - PCSEs don’t solve the problem.
  - Fixed effects, random effects, special effects, can’t get rid of the autocorrelation.

Also, (an old question) how do we choose a level of analysis?

- Do we cut out a wealth of information and study aggregate time series?
  - Many defenders of this: Kramer (1983); MES (1989).
  - This is one way to solve autocorrelation problems – we know how to deal with it at the aggregate level.
- Or, do we ignore dynamics?
  - Throw everyone together and use cross-sectional techniques.
  - Or use PCSTS methods that allow clustering of data without estimation of parameters at the aggregate level.
- Let's do both aggregate- and individual-level.

A Multi-Level-Model using Autoregressive Fractionally Integrated Moving Average techniques on Repeated Cross Sectional Data

- Or, MLM-ARFIMA for RCS
- Think about level-1 units (e.g., people) situated in level-2 structures (e.g., days/months/years).
- MLMs have been used in PCSTS (Beck and Katz 2007; Beck 2007; Shor, Bafumi, Keele and Park 2007). But these solutions either ignore autocorrelation or attempt to fix it with a lagged dependent variable (which we don't have in RCS at the individual-level).
- The MLM relies on the assumption that errors are both spatially and temporally independent. So have to deal with autocorrelation *first*.
- Our solution works for PCSTS but is especially useful for RCS and has less competition there than it does in the PCSTS toolkit.

Key Aspects

- Individual observations are embedded within multiple, sequential time-points.
- Retrieve estimates at the individual-level *and* at the aggregate level.
- Allows use of variables that vary only *within* cross-sections and some that vary between cross-sections (e.g., unemployment rate).
- Box-Jenkins and fractional differencing techniques can control for autocorrelation at level-2. (Box and Jenkins 1976; Box-Steffensmeier and Smith 1996, 1998; Lebo, Walker and Clarke 2000; Clarke and Lebo 2003).
- Introduce *Double Filtering* to clean up two kinds of autocorrelation.

Double Filtering - The Math

Begin with level-2 (aggregate):

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

The ARFIMA equation for long-t time series.

p autoregressive parameters

q moving average parameters

difference d times.

Estimate to find correct values of (p,d,q) to create white noise residuals (ε_t).

Simplifying (1)

Where d=1, model simplifies to: $\Delta\bar{Y}_t = \frac{(1-\theta_q L^q)}{(1-\phi_p L^p)}\varepsilon_t$. (ARIMA)

Where d=0, model simplifies to: $\bar{Y}_t = \frac{(1-\theta_q L^q)}{(1-\phi_p L^p)}\varepsilon_t$. (ARMA)

Choose one based on stationarity tests and direct estimation of d.

More cross-sections will allow better estimate of d.

We should prefer ARFIMA since DGP of series are likely fractionally integrated, even if we don't have data to prove it.

Where t is short, getting d is difficult and ARMA and ARIMA are just approximations of ARFIMA.

First filter: make a noise model for level-2

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

$\overline{Y}_t^*$ is just the residuals from $\overline{Y}_t$ 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.

Next, do this for X_ts and for any variables, Z_t, that vary over time but not within time-points.

With $\overline{Y}_t^*$, $\overline{X}_t^*$, and Z_t^*, level-2 is cleansed of autocorrelation.

Second Filter: cleanse level-1 data

Subtract the daily deterministic component from the level-1 dependent variable:

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

For level-1 covariates, mean-centering (a common practice in PCSTS and MLM)*:

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

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

* 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, $x_{it}^{**} = x_{it} - (\bar{X}_t - \bar{X}_t^{*}) - \bar{X}_t^{*} = x_{it} - \bar{X}_t$.

Now we can estimate the MLM

- Two level equation.
- Level-2 equation can include covariates that vary only between days and those that vary within and between.
  - $\bar{Y}_t^* = \alpha_2 + \beta_2 \bar{X}_t^* + \gamma Z_t^* + u_{2t}$
- The level-1 equation provides the model of within variation:
  - $y_{it}^{**} = \alpha_1 + \beta_1 x_{it}^{**} + u_{1it}$
- Or, can be estimated together, combining both the within and between day effects:
  - $y_{it}^{**} = \alpha_1 + \beta_1 x_{it}^{**} + u_{1it} + \beta_2 \bar{X}_t^* + \gamma Z_t^* + u_{2t}$

Points of Flexibility

- Depending upon the length of T, one might estimate an ARMA, ARIMA, or ARFIMA model to create an appropriate noise model. E.g., an AR(1) might be best for cumulative NES.
- If (0,0,0) the model reduces to mean centering.
- Can include time varying coefficients and inclusion of level-1 data, W_{it}, not in all waves: $Y_{it}^{**} = \alpha_{1t} + \beta_{1t} X_{it}^{**} + \delta_t W_{it} + u_1$.
- Applicable to PCSTS, but more tools there.
  - PCSE, Differencing, Lags, etc.

Monte Carlo Analyses

- We know the added value of estimating cross-sectional and dynamic parameters together.
- But has double filtering solved the two directions of autocorrelation?
- We expect that the greater the time dependence in level-2, the greater the degree of bias in coefficients.
- If observations at time t are more correlated with one another than with observations at $t+s$, this is a problem of clustering in the data; the errors will not be independent and the standard errors will be incorrect.

Monte Carlo Setup

- We generate level-2 time series with varying levels of memory (serial correlation).
- Each aggregate value gives us a mean for a distribution from which to draw individual-level data.
- We do this for *X*s and *Y*s.
- We also generate data with no serial correlation to serve as a baseline for comparison.

Monte Carlo Comparisons

- We test the statistical properties of eight approaches:
  - OLS (naïve) pooling all the data
  - OLS separating between and within day effects (OLS)
  - OLS with day-level lag (OLS-LDV)
  - OLS single filtering (OLS-ARFIMA) without level-2
  - Multi-level model (naïve) with time-varying intercepts
  - MLM separating between and within day effects (MLM)
  - MLM with day-level lag (MLM-LDV)
  - Our double filtering method (MLM-ARFIMA)

Simulation Expectations I - OLS

- Naïve pooling should lead to bias and inefficiency since it ignores clustering.
- A lagged dependent variable, $\bar{Y}_{t-1}$, will lead to unbiased and efficient estimates if it's enough to clean up the autocorrelation at level-2. (But there is a ton of literature showing that it never is)
- ARFIMA methods should minimize bias and inefficiency.

Simulation Expectations II - MLM

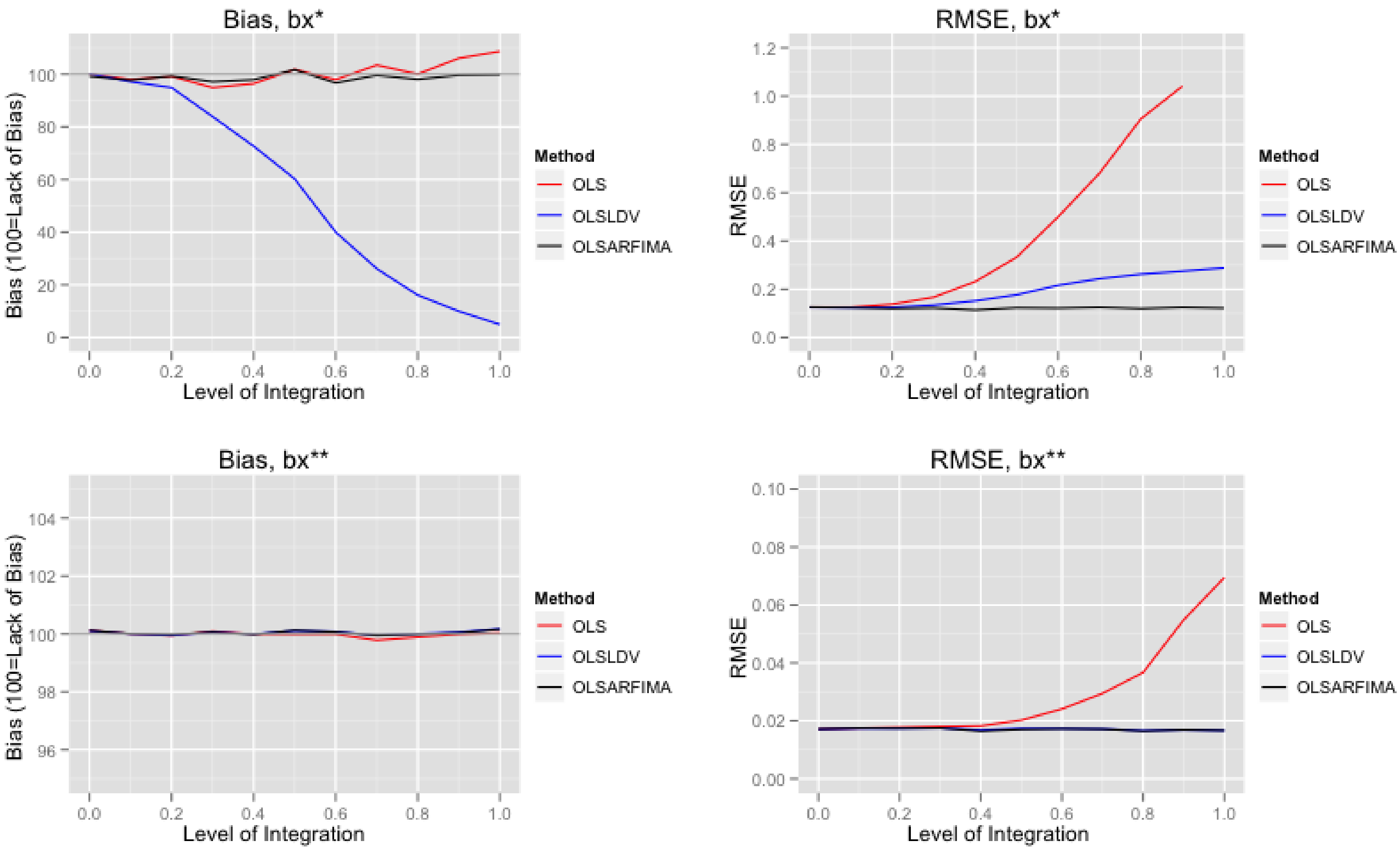
- The MLM approaches should be an improvement over OLS by accounting for the clustering in the data.
- However, an assumption of the MLM is that level-2 errors will be independently distributed, which is violated insofar as ARFIMA properties are unaccounted for at level-2.
- Simple MLM, will produce biased and inefficient estimates as *d* increases.
- Similarly, MLM-LDV – the multilevel model with a level-2 lagged dependent variable -- will produce estimates that are biased downward as *d* increases. This occurs as the level-1 units are not filtered at all.
- MLM-ARFIMA should fix everything, we hope.

Simulation Results

- We calculate comparative bias as $(\bar{\theta} - \theta)/\theta$ where $\bar{\theta}$ and θ are the estimated and true values, respectively.
- RMSE= $\sqrt{\frac{\sum(\bar{\theta}-\theta)^2}{n}}$, where n is the number of replications in each cell (i.e., 1,000). A small RMSE is preferred over a large RMSE, as it indicates less variation around the true population value.
- The standard errors calibrated to the size of the sample variation, which is "optimism" or "overconfidence" (Beck and Katz 1995; Shore et al 2007).
- $Optimism = 100 \times \sqrt{\frac{\sum_{l=1}^{1000}(\beta_l-\bar{\beta})^2}{\sum_{l=1}^{1000} SE\beta_l}}$

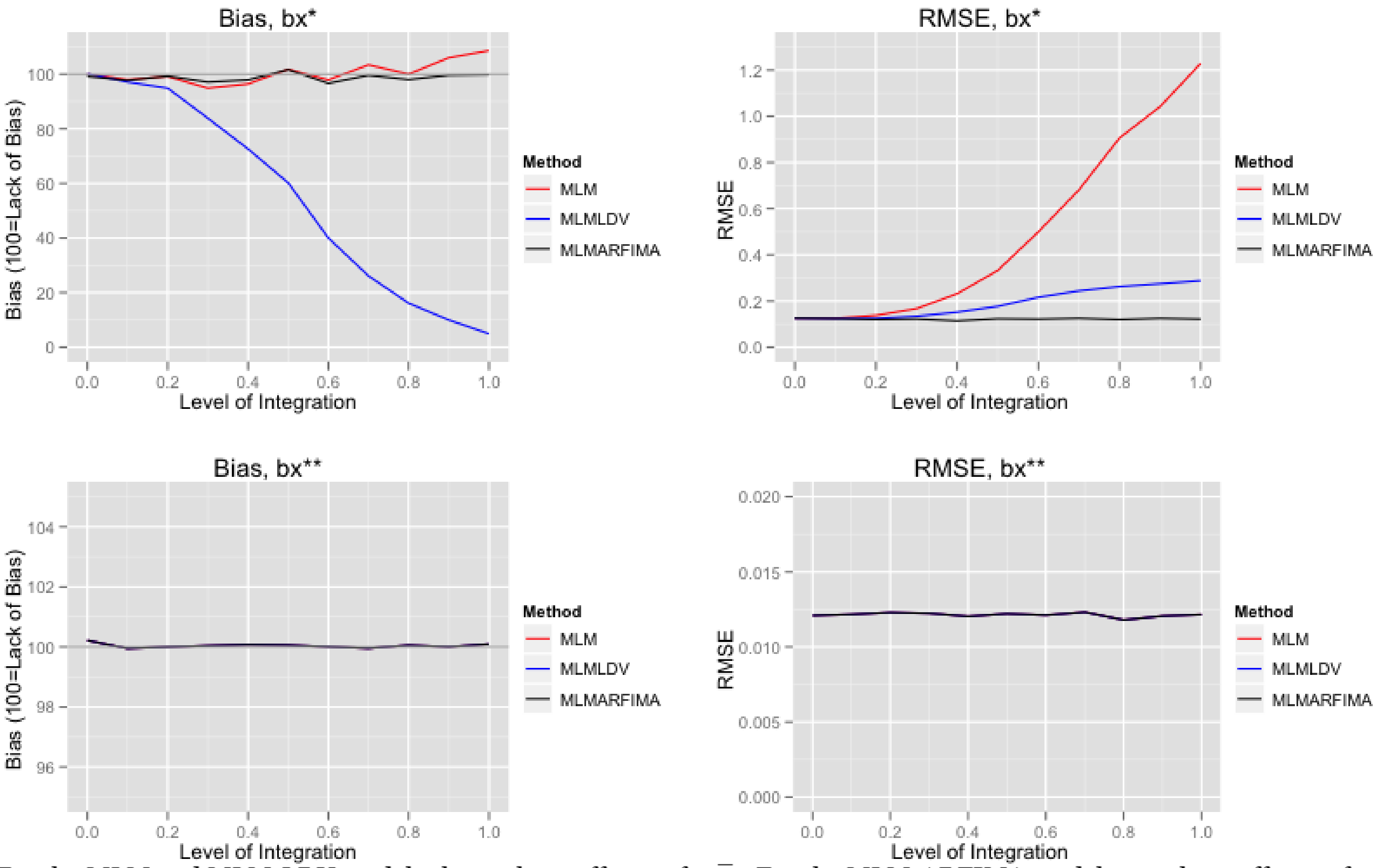
Over 100 means SE are smaller than they should be – we are too confident.

Bias and RMSE for OLS Coefficients*



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

Bias and RMSE for Multilevel Models*



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

Optimism Index

	Between Day Effects (bx*)		
d	OLS	OLSLDV	OLS-ARFIMA
0	747	1818	734
0.1	747	1695	723
0.2	818	1601	702
0.3	1025	1346	713
0.4	1462	1101	666
0.5	2189	1007	721
0.6	3432	1419	709
0.7	4875	2283	730
0.8	6525	3636	702
0.9	7480	5172	732
1.0	8344	7231	711

Optimism Index

	Between Day Effects (bx*)		
d	MLM	MLM-LDV	MLM-ARFIMA
0	105	255	103
0.1	104	238	101
0.2	113	224	99
0.3	138	188	100
0.4	190	153	94
0.5	269	139	101
0.6	396	196	100
0.7	534	316	103
0.8	685	506	99
0.9	766	723	103
1.0	843	1011	100

	Optimism βx^*	Optimism βx^{**}	Bias, βx^*	RMSE, βx^*	Bias, βx^{**}	RMSE, βx^{**}
OLS	X	☺	X	X	☺	X
OLS-LDV	X	☺	X	X	☺	☺
OLS-ARFIMA	X	☺	☺	☺	☺	☺
MLM	X	☺	X	X	☺	☺
MLM-LDV	X	☺	X	X	☺	☺
MLM-ARFIMA	☺	☺	☺	☺	☺	☺

Application: The 2008 Presidential Election

- To what extent did the economy favor Obama? (Kenski, Hardy and Jamieson 2010)
- Scholarly and conventional wisdom=poor economic circumstances favor the Democratic candidate (Kenski, Hardy and Jamieson 2010).

$$y_{it}^{**} = \alpha_1 + \beta_1 x_{1it}^{**} + \beta_1 x_{2it}^{**} u_{1it} + \cdots \beta_j x_{jit}^{**} +$$
$$+\omega_1 \bar{X}_{1t}^{**} + \omega_1 \bar{X}_{2t}^{**} + \cdots \omega_k \bar{X}_{1k}^{**} +$$
$$+\gamma_1 Z_{1t}^{**} + \gamma_1 Z_{2t}^{**} + \cdots \gamma_k Z_{1k}^{**} +$$
$$u_{1it} + u_{2t}$$

Application: The 2008 Presidential Election

Individual-level predictors

$$y_{it}^{**} = \alpha_1 + \beta_1 x_{1it}^{**} + \beta_1 x_{2it}^{**} u_{1it} + \cdots \beta_j x_{jit}^{**}$$

PID, income, age, gender……

Aggregate Level Survey Predictors

$$+\omega_1 \bar{X}_{1t}^{**} + \omega_1 \bar{X}_{2t}^{**} + \cdots \omega_k \bar{X}_{1k}^{**} +$$

Aggregate PID, income……

Aggregate Level Predictors

$$+\gamma_1 Z_{1t}^{**} + \gamma_1 Z_{2t}^{**} + \cdots \gamma_k Z_{1k}^{**} +$$

Economic conditions

Day Level Error

Within Day Error

$$+u_{1it} + u_{2t}$$

Measures

- DV: Comparative Evaluation (Evaluation Obama-Evaluation McCain)
- Individual level predictors (x^{**}): PID, economic evaluation, income, age, gender.
- Aggregate survey predictors (X^{*}): economic evaluation, income, PID.
- Aggregate predictors (Z^{*}): DJIA

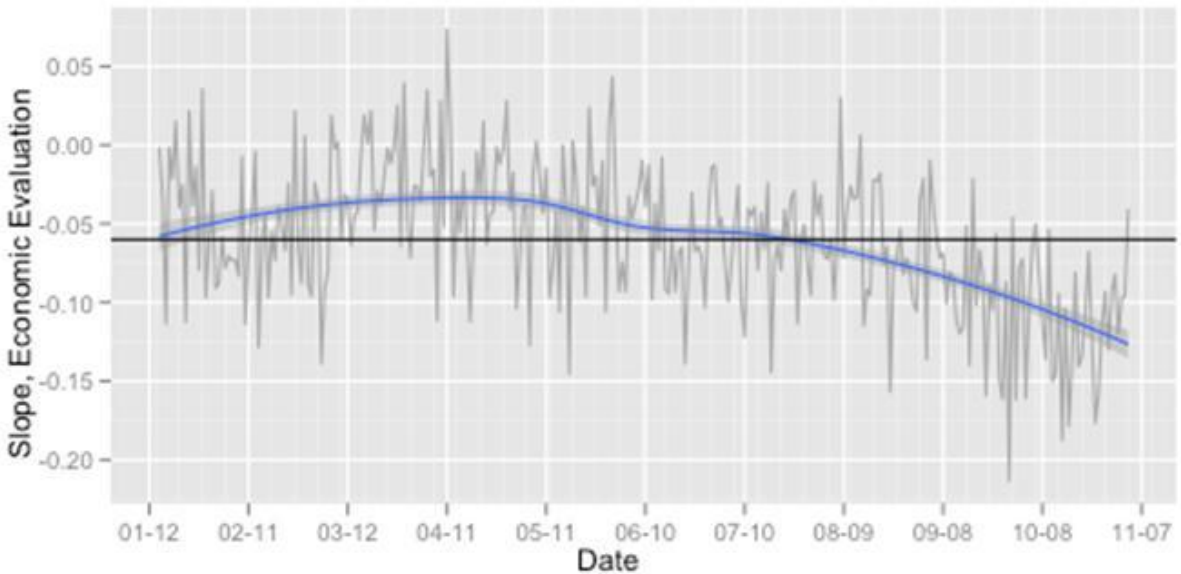
	OLS-Naive	*OLS*	*OLS-LDV*	*OLS-FI*	*MLM-Naive*	*MLM*	*MLM-LDV*	*MLM-FI*
<u>Between Effects</u>								
Evaluation	---	**-0.334**	**-0.288**	0.057	---	**-0.330**	**-0.285**	0.050
		(.1049)	**(.1052)**	(.158)		**(.1340)**	**(.1300)**	(.190)
PID (Democrat)	---	**1.312**	**1.240**	**1.256**	---	**1.277**	**1.212**	1.233
		(.0972)	**(.0979)**	**(.0980)**		**(.1231)**	**(.1200)**	(.1171)
Personal Income	---	0.032	0.110	0.120	---	-0.017	0.068	0.090
		(.2127)	(.2131)	(.208)		(.270)	(.2618)	(.2491)
DJIA	**-0.004**	**-0.006**	**-0.005**	0.008	**-0.004**	**-0.006**	**-0.005**	0.008
	(.0014)	**(.0016)**	**(.0016)**	(.008)	**(.0014)**	**(.0021)**	**(.0020)**	(.01)
Lag Y			**0.242**				**0.236**	
			(.0417)				**(.0520)**	
Intercept	**-4.373**	**-4.022**	**-4.200**	**-5.42**	**-4.373**	**-3.670**	**-3.910**	**-5.218**
	(.1929)	**(1.031)**	**(1.031)**	**(.0899)**	**(.1929)**	**(1.313)**	**(1.272)**	**(1.076)**

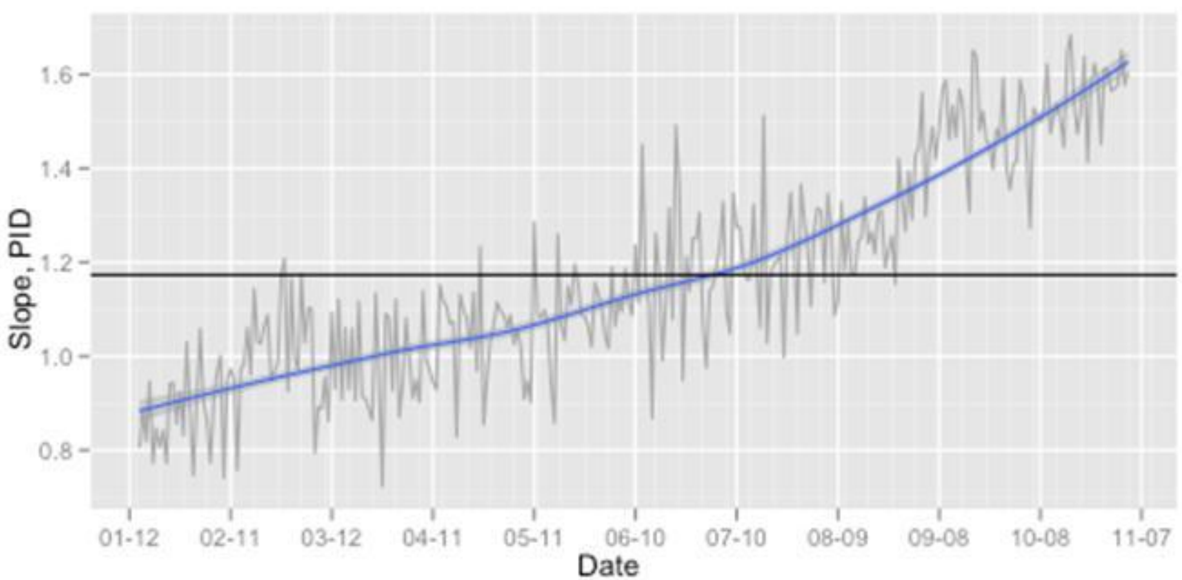
Within Effects	*OLS-Naïve*	*OLS*	*OLS-LDV*	*OLS-FI*	*MLM-Naive*	*MLM*	*MLM-LDV*	*MLM-FI*
Evaluation	-0.069	-0.064	-0.064	-0.064	-0.067	-0.064	-0.064	-0.064
	(.0139)	(.014)	(.0140)	(.0140)	(.0139)	(.014)	(.0140)	(.0140)
PID (Democrat)	1.200	1.200	1.200	1.200	1.200	1.200	1.200	1.200
	(.0074)	(.0074)	(.0074)	(.0074)	(.0074)	(.0074)	(.0074)	(.0074)
Personal Income	-0.053	-0.055	-0.055	-0.055	-0.054	-0.055	-0.055	-0.055
	(.0196)	(.0197)	(.0197)	(.0197)	(.0196)	(.0197)	(.0197)	(.0197)
Age	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Female	0.245	0.246	0.246	0.246	0.245	0.246	0.246	0.246
	(.033)	(.033)	(.033)	(.033)	(.033)	(.033)	(.033)	(.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

Application Results and Extension

- Real economic conditions did not affect evaluations, controlling for non-stationarity.
- Aggregate economic evaluations also did not impact evaluations, controlling for non-stationarity.
- Dynamic, day-level effects.
- We specified a random slope for the within-day coefficient associated with PID and economic evaluations


Slope, Economic Evaluation
0.05
0.00
-0.05
-0.10
-0.15
-0.20
01-12
02-11
03-12
04-11
05-11
06-10
07-10
08-09
09-08
10-08
11-07
Date


Slope, PID
1.6
1.4
1.2
1.0
0.8
01-12
02-11
03-12
04-11
05-11
06-10
07-10
08-09
09-08
10-08
11-07
Date

Application Results and Extension

- MLM-ARFIMA is a relatively flexible model
- Solves problems of bias and inefficiency
- Accounts for unobserved heterogeneity and dynamic effects in RCS data
- Gets more out of the data
- Easily extended to more complex hierarchical designs
- Future directions
  - Apply to true panels and test against panel methods
  - Dichotomous and Likert dependent variables.
  - Applications