Choosing and Combining Units: Common Problems in Multilevel and Cross-Temporal Research

Laura Stoker
Department of Political Science
University of California, Berkeley

ESRC 2012 Summer Methods Festival
Oxford University

I. Macro-level Hypotheses with Macro & Micro-level Data Generating Processes

The limitations of testing macro-level hypotheses with macro-level data.

The broader question of understanding dynamics across levels of analysis—macro to micro and micro to macro.

II. Choosing Temporal Units of Analysis: The Issue of Granularity

How fine-grained a unit should one select for a time-varying Y—minute, hour, day, week, year, decade?

III. Missed opportunities for Temporal Disaggregation

Longitudinal analysis of a single cross-sectional sample survey; Difference in Difference estimation with or without temporal aggregation

I. Macro-level Hypotheses—Some Examples

1. Xj: Electronic vs. Paper & Pencil Records Yj: Clinic Efficiency (# Patients seen/day)
   J: Clinic

2. Xj: Open Seat vs. Incumbent Running Yj: Closeness of Election
   J: U.S. Congressional District

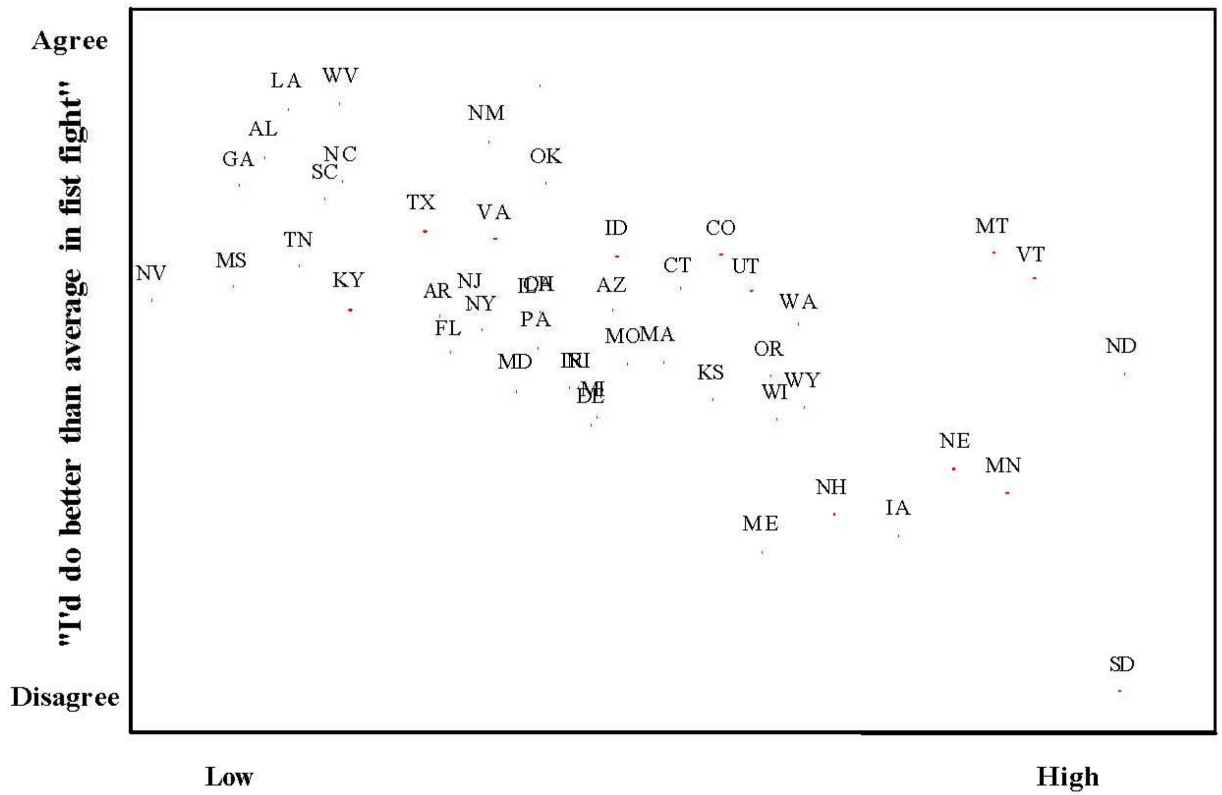
3. Xj: Ease of Registration to Vote Yj: State-level Turnout
   J: U.S. State

4. Xj: Social Capital (% Trusting & Joining) Yj: Rate of Crime, Mortality, Tolerance….
   J: U.S. State

5. Xj: % Catholic in State Yj: State-level Abortion Laws
   J: State

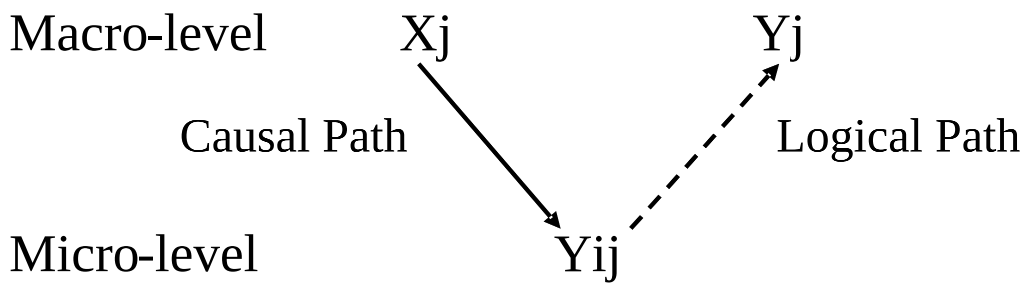
6. Xj: Seniority Yj: Legislative Success (% bills that pass)
   J: Member of Legislature

Figure 7.5
States high in social capital are less pugnacious



Source: "Social Capital: Measurement and Consequences," by Robert Putnam.

In such cases, a stylized [and, to be sure, simplified and reductionist] representation of the data generating process looks like this:



Some analysts work entirely with macro-level data, estimating this relationship directly:

$$Xj \longrightarrow Yj$$

Other analysts focus on the left-hand side but ignore the right-hand side.

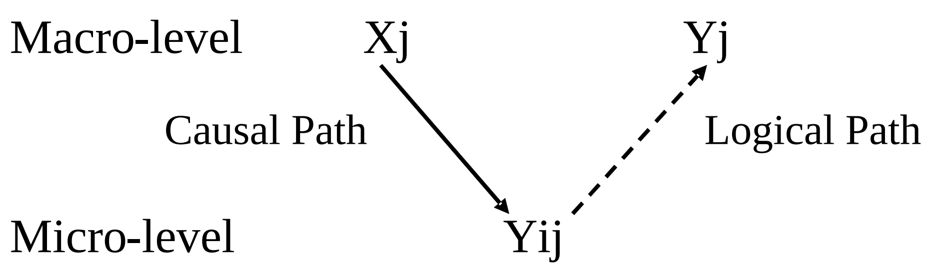
So What?

A. Because an analyst in the first group is interested in $Xj \rightarrow Yj$, and (let's say), has no interest at all in $Xj \rightarrow Yij$, he or she is *not guilty* of the ecological fallacy in analyzing J-level data.

B. Yet, for reasons that I will lay out below, the better practice is to:

1. Estimate the effect of Xj on Yij in a multi-level framework.

2. Then use logical induction to determine the effect of Xj on Y



C. Analysts in the second group miss an opportunity to develop the macro-level implications of micro-level dynamics (themselves tied to macro-level developments).

X: Electronic vs. Paper & Pencil Records	
Yij: Doctor Efficiency (# Patients seen by doctor/day)	I: Doctor
Yj: Clinic Efficiency (# Patients seen/day)	J: Clinic
Xj: Open Seat vs. Incumbent Running	
Yij: Individual Vote Choice, Individual Turnout	I: Citizen
Yj: Closeness of Election	J: Congressional district
Xj: Ease of Registration to Vote	
Yij: Individual Turnout	I: Citizen
Yj: State-level Turnout	J: State
Xj: Social Capital Index	
Yij: Individual Pugnacity	I: Citizen
Yj: State Pugnacity	J: State
Xj: % Catholic in State	
Yij: Votes on Abortion Laws	I: Legislator
Yj: State-level Abortion Laws	J: State
Xj: Seniority	
Yij: Did bill sponsored pass?	I: Bill
Yj: Legislative Success (% bills that pass)	J: Legislator

To start, suppose a very simple random intercepts multilevel model:

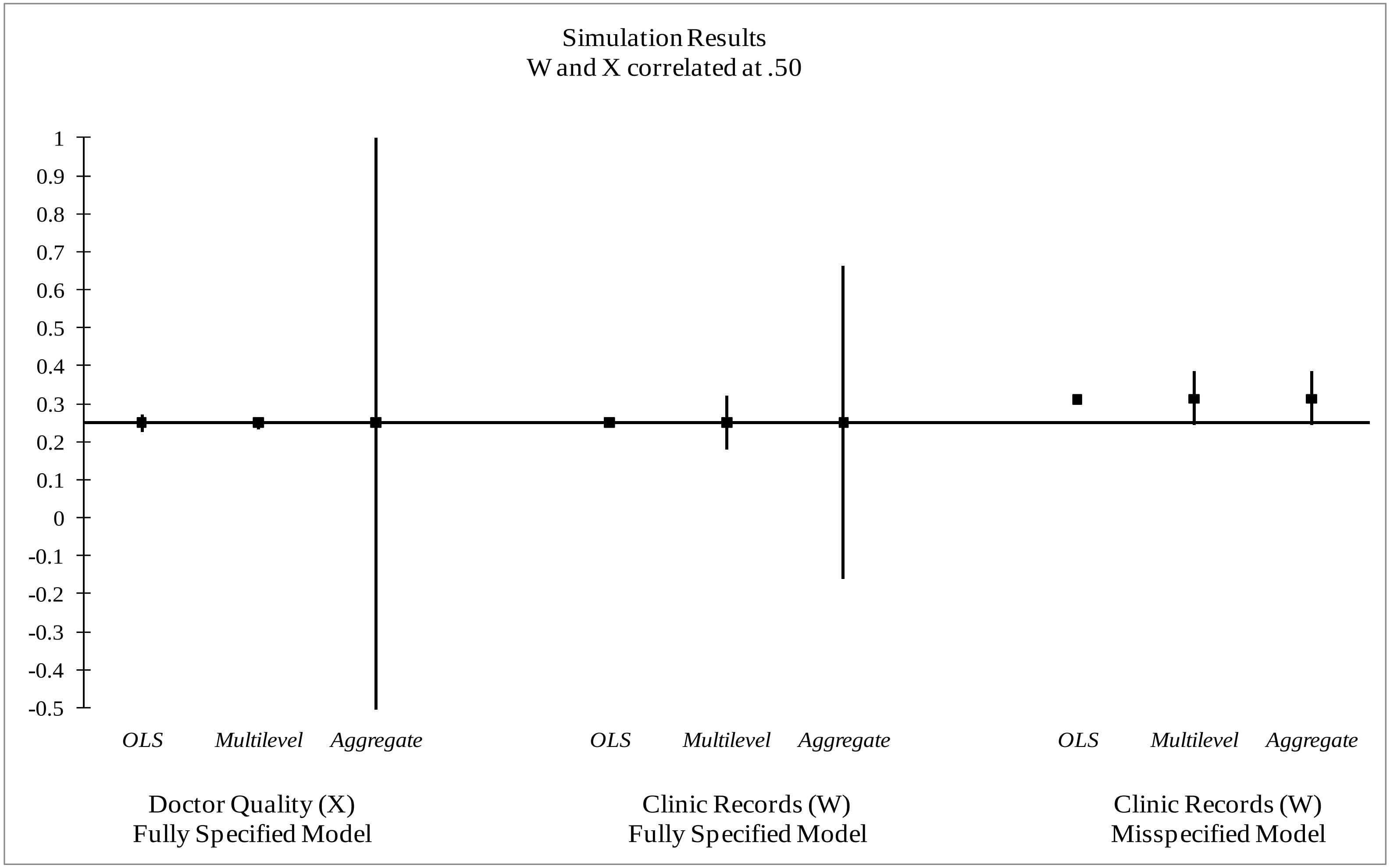
Level 1 (doctor):	$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + e_{ij}$	(1)
Level 2 (clinic):	$\beta_{0j} = \gamma_{00} + \gamma_{01} w_{1j} + u_{0j}$	(2)
Reduced Form:	$y_{ij} = \gamma_{00} + \gamma_{01} w_{1j} + \beta_1 x_{1ij} + e_{ij} + u_{0j}$	(3)

y_{ij} — the typical number of patients the ith doctor in the jth clinic sees in a day.
x_{1ij} — “doctor quality” (say), ranging from low (0) to high (1) quality.
w_{1j} — clinic system of record-keeping system, paper & pencil (0) or electronic (1).
e_{ij} — doctor-level error term
u_{0j} — clinic-level error term

Suppose you regress $\bar{y}_j$ on $\bar{x}_{1j}$ and w_{1j} instead of regressing y_{ij} on x_{1ij} and w_{1j}.

Problem #1: Inefficiency

So long as some of the variation in x_{1ij} is between-group variation, the correlation between $\bar{x}_{1j}$ and w_{1j} will be higher in absolute value than the correlation between x_{1ij} and w_{1j}. Estimation will be less precise.



True: $y_{ij} = .25w_{1j} + .25x_{1ij} + e_{ij} + u_{0j}$, e_{ij} and u_{0j} iid normal (0, .0625), n =10,000, J =100, 50 $\leq n_j \leq$ 150
Shown are average coefficients and standard errors across 1,000 simulations. Results labeled "OLS" and "Multilevel" use the disaggregated data. The "aggregate" analysis gives results from an OLS regression of $\bar{y}_{ij}$ on w_{1j} and $\bar{x}_{1j}$

PROBLEM #2: BIASED ESTIMATES OF THE EFFECT OF w_{1j}

A. Ignoring the individual-level process and neglecting to include (aggregated) individual-level variables.

B. Bias tied to Problems involving Non-linear or Non-additive x_{ij} effects

If the effects of the individual-level Xs are non-linear or non-additive, analysis at the aggregate data will typically produce biased estimates of those effects.

> Non-linear (e.g., quadratic):
> How y_{ij} varies with $x_{1ij} * x_{1ij}$ across individuals will in most circumstances not correspond to how $\bar{y}_j$ varies with $\bar{x}_{1j} * \bar{x}_{1j}$ across groups.
>
> Non-additive (e.g., product interaction):
> How y_{ij} varies with $x_{1ij} * x_{2ij}$ across individuals will not ordinarily equal how $\bar{y}_j$ varies with $\bar{x}_{1j} * \bar{x}_{2j}$ across groups.

This is well known from the literature on ecological inference and the ecological fallacy. The reason we care about this is because bias in those estimates will, in turn, bias estimates of the effect of the macro-level independent variable (w_{1j}).

C. Bias in aggregate-level estimation of what are actually cross-level interactions.

The analyst is (or could be) interested in how the effect of the macro-level variable depends on individual-level characteristics. Of interest, here, is the effect of $x_{1ij} * w_{1j}$ on y_{ij} but the aggregate-level analyst is considering how $\bar{x}_{1j} * w_{1j}$ affects $\bar{y}_j$. In many circumstances, the two will not agree.

D. The model excludes important x_{ij} variables that are correlated with w_{1j}.

Bias due to e_{ij} (and u_j) correlated with w_{1j} will bedevil a multi-level analysis as well, but becomes more severe when the analysis is aggregated:

"The cause is the averaging process that aggregation imposes on data. After averaging, much individual variation is smoothed out, and central limit theorems have their usual effects. . . . . The result is that biases too small to be of great concern in individual-level data become powerful sources of bias in the aggregate data. An imperfect but serviceable specification at the microlevel becomes useless at the macrolevel." (Achen and Shively 1995, p. 110)

E. The macro-level variable of interest is an aggregate variable — $\bar{x}_{1j}$ *— and* y_{ij} *is affected by both* $\bar{x}_{1j}$ *and* x_{1ij}

Examples:

As own income ↑, conservatism ↑
As average income in state ↑, conservatism ↓

As own time-to-diagnosis in a hospital ↑, the probability of having unnecessary surgery ↑
As average time-to-diagnosis in hospital ↑, the probability of having unnecessary surgery ↓

The aggregate-level analysis is essentially giving you the average of the two effects. Unless the two effects are the same, the aggregate analysis will get both of them wrong.

F. The macro-level variable of interest is a relational variable —one that targets some micro-level units and not (or more than) other micro-level units.

Examples:

Campaign to mobilize Democrats (but not Republicans) within a place
Tax policy that benefits those earning > $100k but not those earning < $100k

Two ways to think about relational macro-level variables:

(1) Relational macro-level variables are effectively individual-level variables, e.g.:

Score 0 if individual lives in place without tax policy
Score 1 if individual lives in place with tax policy and is not targeted (make < $100k)
Score 2 if individual lives in place with tax policy and is targeted (make > $100k)

(2) Hypotheses about relational macro-level variables require cross-level interactions—i.e., in this example, the hypothesis implies a model that includes:

w_{1j} whether live in place with the tax policy in question

x_{1ij} whether or not make more than $100k

$w_{1j} * x_{1ij}$

Either way, addressing the macro-level hypothesis requires individual-level data.

Anderson, Box-Steffensmeier, and Sinclair-Chapman (2003)

Key question: What explains legislative success?
The ideas:

i	Legislative initiative ("bill") one has sponsored
j	Member of U.S. Congress (MC)
y_{ij}	Probability that a bill that an MC sponsors passes
$\bar{y}_j$	Legislative Success—proportion of bills that an MC sponsors that pass
x_{1ij}	Speaking vs. not speaking on behalf of a bill one sponsored (sending a cue as to a bill's importance)
$\bar{x}_{1j}$	Being voluble, generally ("delivering too many speeches might result in a member being perceived as difficult or obstructionist")
x_{2ij}	Hot Bills ("that maybe be able to ride a wave of political interest into legislative success")
x_{3ij}	Local Bills
w_{1j}	Seniority of MC
w_{2j}	MC's prior margin of electoral victory
w_{3j}	Party of MC

The analysis is carried out at the MC level. Among other things:

$\bar{x}_{1j}$ represents the effects of $\bar{x}_{1j}$ and x_{1ij}

$\bar{x}_{2ij}$ (% of bills introduced that are hot) and $\bar{x}_{3ij}$ (% of bills introduced that are local) are used, respectively, to estimate the effects of x_{2ij} and x_{3ij}

A concocted (half-baked) example:

Key Question: Do state voter registration laws affect state turnout?

Why state turnout?

(1) Normative—concerns about systematic political inequality
(2) Policy-related—where would a policy change make the most difference?
(3) Theoretical—State-level turnout important to theory of, say, party competition

What I did:

Modeled individual-level turnout as a function of age (recoded to range from 0-1), age-squared, educational attainment (5-point scale), and "state-level registration laws"—1=easy, 0=hard.

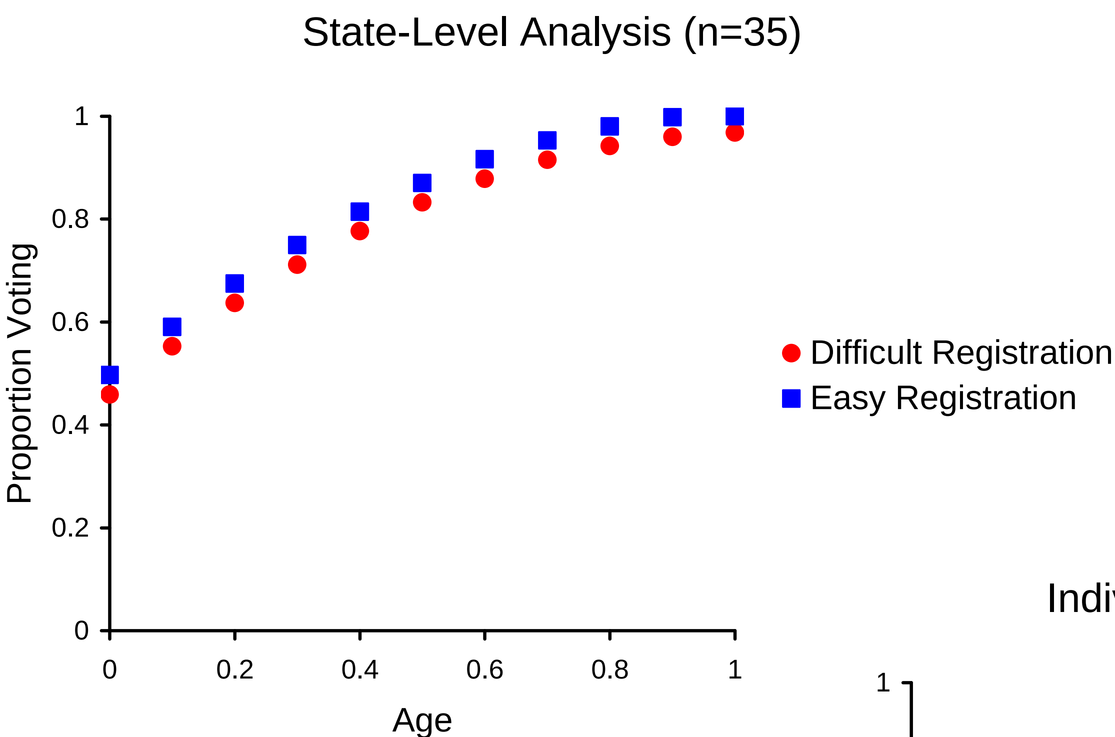
To keep it simple, used linear probability model and OLS.

Better: Get better (real) measure of state registration law variation; use sigmoid probability model and Logit/Probit; match on covariates

Ran individual-level model and also aggregate, state-level model

Goals here:

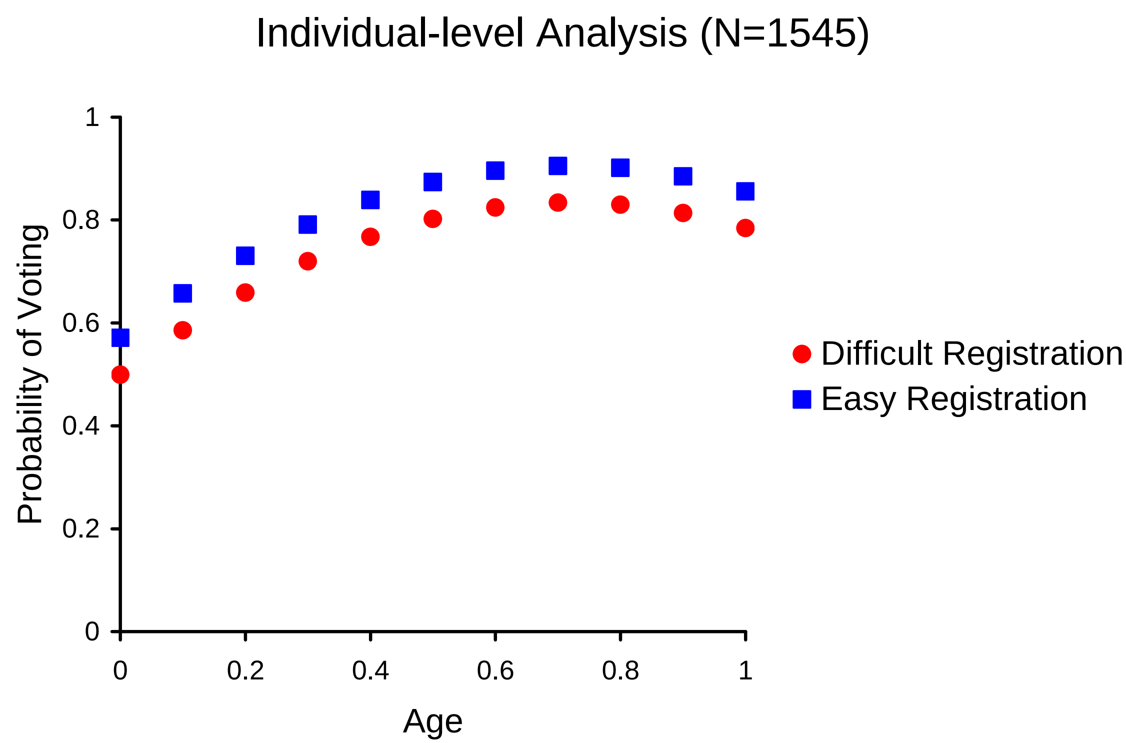
To demonstrate how findings regarding the macro-level hypothesis differ across approaches
To illustrate an induction of macro-level implications from micro-level results.



Registration Law Coefficients

State-level
b=.037, p=.231 (n=35)

Individual-level
b=.074, p=.003 (n=1503)



Moving to the Macro Level:

(1) In states with hard registration, find percentage of state residents who would be turned from non-voters to voters if the policy were to change (to easy).

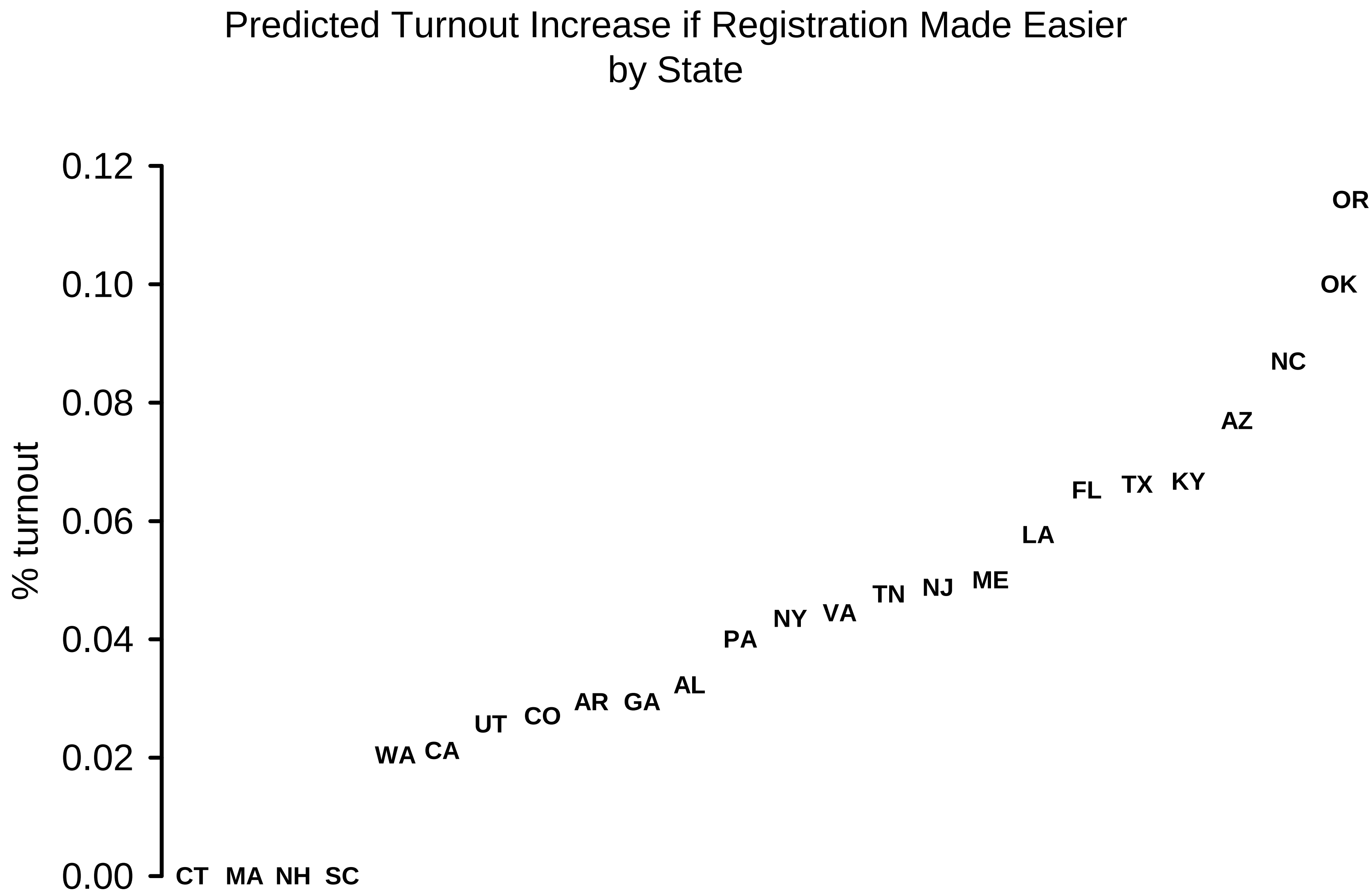
> If probability of voting is < .426, no change—still would not vote
> If probability of voting is > .500, no change—already votes
> If probability of voting is between .426 and .500, would turn from non-voter to voter
>
> % of individuals turning from non-voter to voter: 4.38%[♦]
> Average state-level increase in turnout: 4.40%

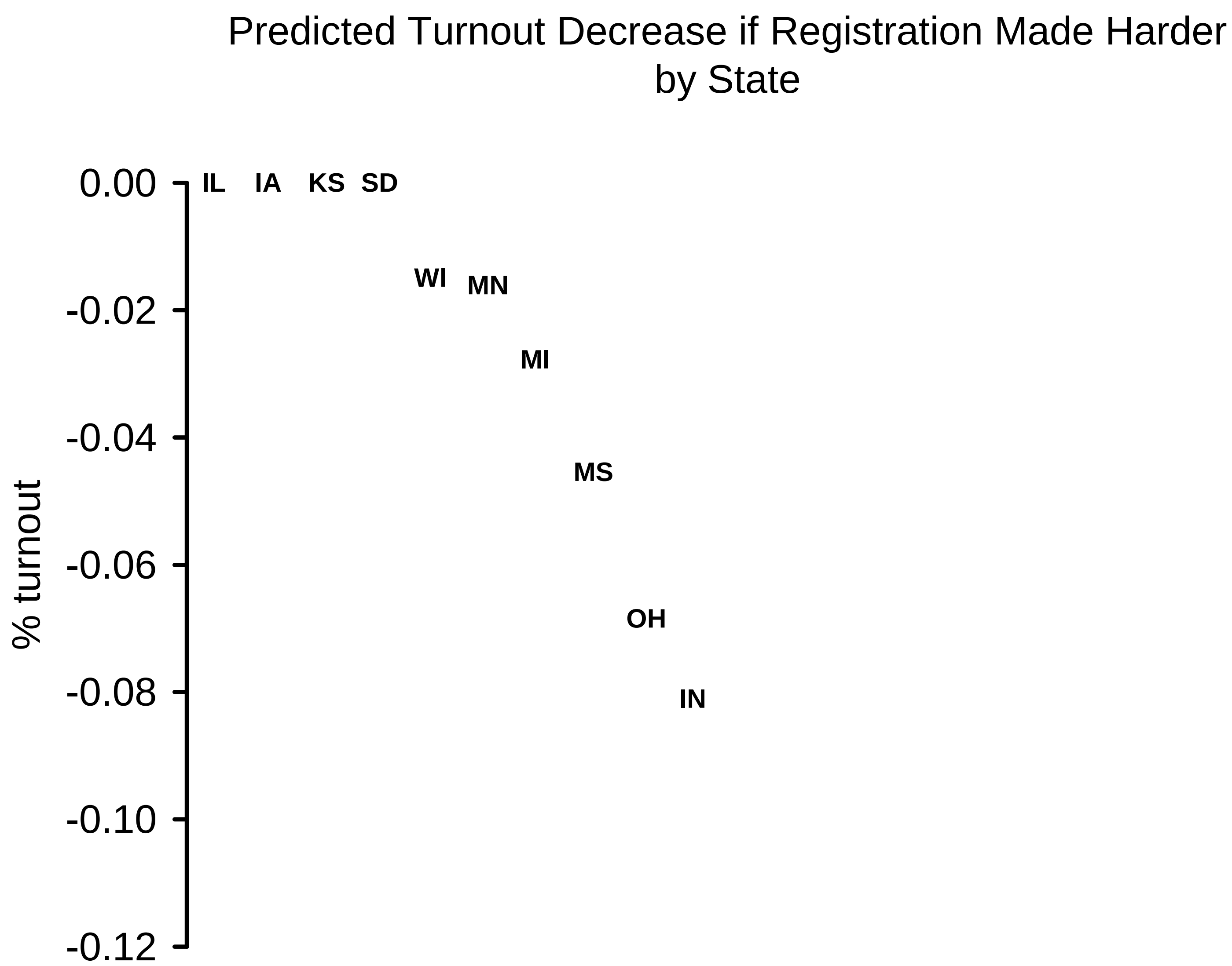
(2) In states with easy registration, find percentage of state residents who would be turned from voters to non-voters if the policy were to change (to hard).

> If probability of voting is >.574, no change—still would vote
> If probability of voting is <.500, no change—already does not vote
> If probability of voting is between .574 and .500, would turn from voter to non-voter
>
> % of individuals turning from voter to non-voter: 2.91%[♣]
> Average state-level decrease in turnout: 2.54%

♦ Analogue of average treatment effect on the untreated

♣ Analogue of average treatment effect on the treated


Predicted Turnout Increase if Registration Made Easier
by State
% turnout
0.12
0.10
0.08
0.06
0.04
0.02
0.00
CT
MA
NH
SC
WA
CA
UT
CO
AR
GA
AL
PA
NY
VA
TN
NJ
ME
LA
FL
TX
KY
AZ
NC
OK
OR


Predicted Turnout Decrease if Registration Made Harder
by State
% turnout
0.00
-0.02
-0.04
-0.06
-0.08
-0.10
-0.12
IL
IA
KS
SD
WI
MN
MI
MS
OH
IN

Moving from Micro-level findings to Macro-level Implications

1. Very simple in some instances—e.g., doctor patients/day to clinic patients/day

2. Manageably simple in other instances—e.g., from the probability that an individual will vote to the proportion in a given place that vote.

3. But not simple/manageably simple if we abandon the atomistic view of "individual units"—e.g., bring in diffusion, contagion, thresholds, and so on.

4. And not simple/manageably simple if we add complexity to the theoretical statement of how macro-level Xs affect macro-level Ys. Even simple elaborations can make the logical induction from micro to macro quite difficult. For example, if a targeted political campaign (macro X) affects both the turnout (Y1) and the vote choice (Y2) of the citizens it reaches.

5. And, of course, the Micro→Macro link may not (in many cases) be logical, but instead be causal. This is at least as worthy of investigation as the kind of case considered here.

6. For at least some problems, the better strategy may be to build a complex (e.g., agent-based) model and develop testable implications that can be evaluated with simpler research designs and data collections.

To Sum Up:

Macro-level hypotheses that involve macro & micro-level data generating processes should be evaluated through analyses that work with data on both (or all) levels even if the proposition of interest is only directly concerned with the macro level.

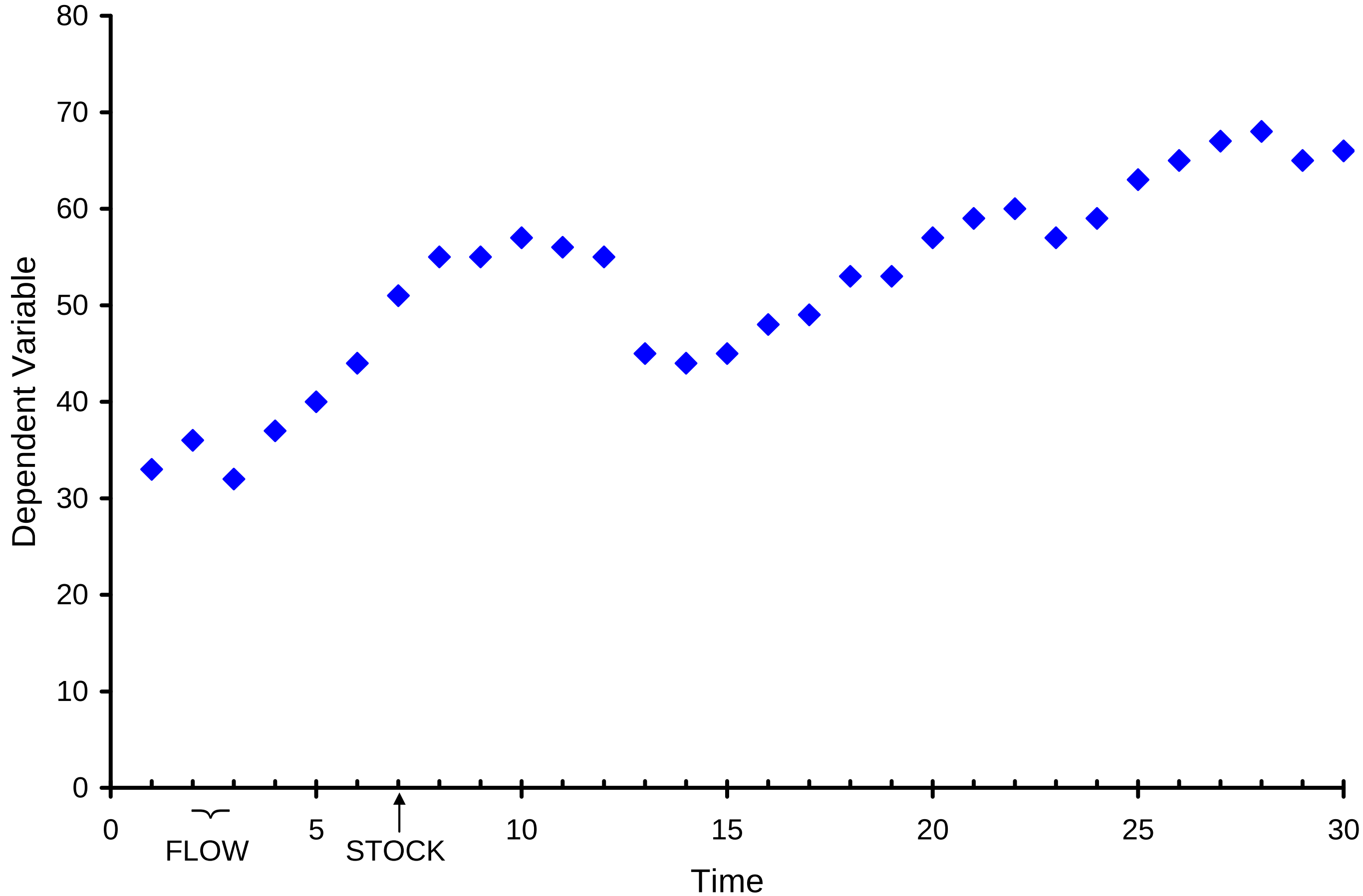
Working with (observational) macro-level data, alone, produces inefficiencies, bias in estimation, and is just-plain incapable of representing the dynamics involved.

1. Inefficient, even when no bias introduced (e.g., macro-X is experimental)
2A. Bias: neglect individual-level variables altogether
2B. Bias: Xij effects are non-linear or non-additive
2C. Bias: using aggregate version of cross-level interaction
2D. Bias: magnified consequences of errors correlated with Xij (and, in turn Wi)
2E. Bias: both Xij and X-bar affect Yij
2F. Bias: Macro-level X is relational

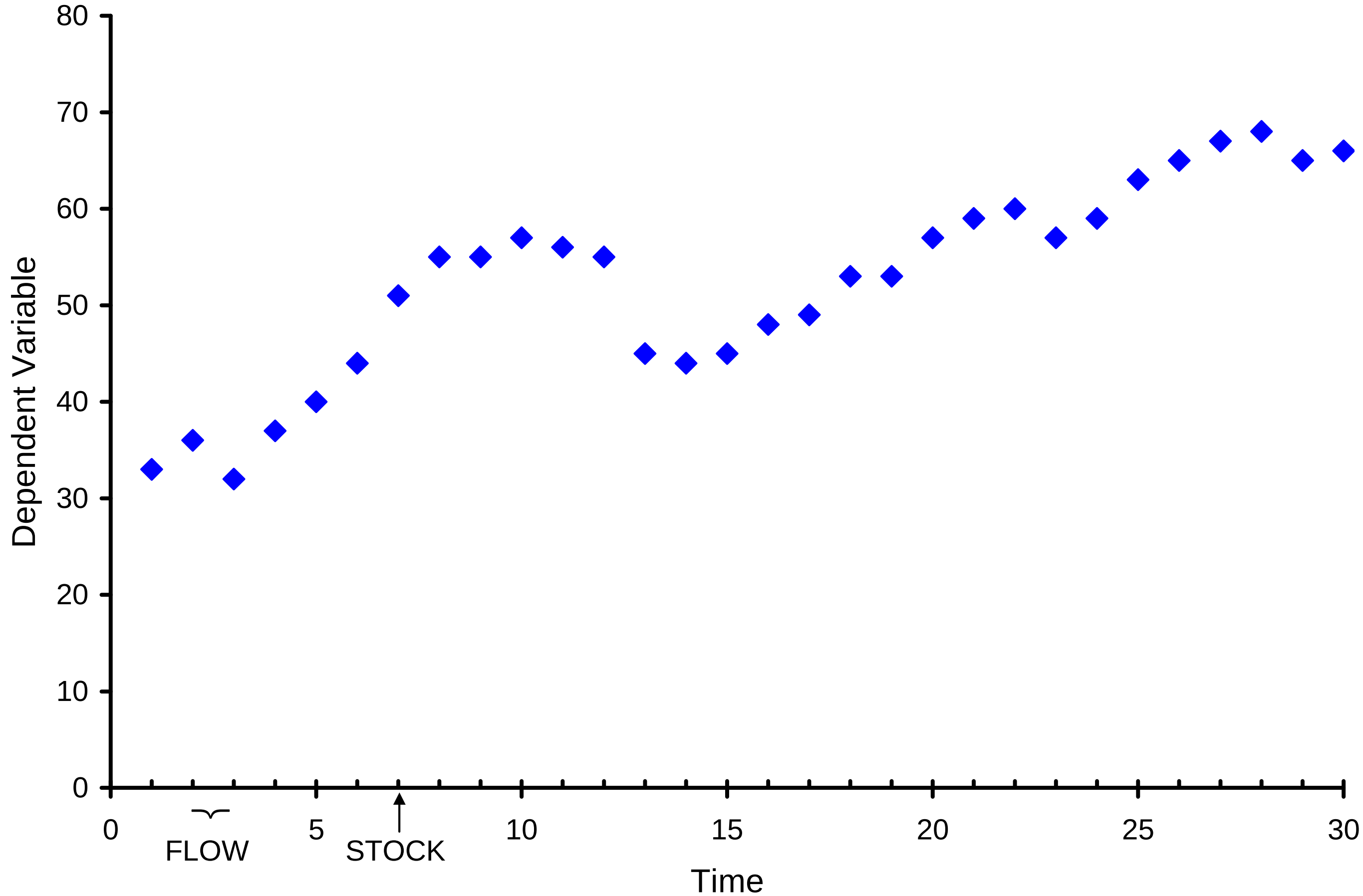
The advice:

- Evaluate how macro-level Xs affect micro-level Ys even if you have no interest whatsoever in "political behavior."
- Use logical induction to move from micro-level results to macro-level implications—the key move that must be made in order to test the original macro-level proposition.
- Go well beyond the simple (and reductionist) set-up I have worked with here. The ultimate goal is a more sophisticated and rich understanding of cross-level dynamics.

II. Choosing Temporal Units of Analysis: The Issue of Granularity



- <u>Stocks</u>: measured at points in time (e.g. stockpile of weapon, prices)
- <u>Flows</u>: measured across periods of time (e.g. arms acquisition over a period, events—as discussed later)



- Problem: Results are sensitive to choice of temporal units/intervals
- Giant literature on this—in economics, especially. Common advice found in my field is to evaluate sensitivity of results to the choice of temporal unit.
- Question: Is there a “Natural unit” (stocks) or “natural interval” (flows)?

- Natural unit (NTU) of a stock: "The shortest unit that satisfies the condition that all causally meaningful variation is *between* and *not within* units."

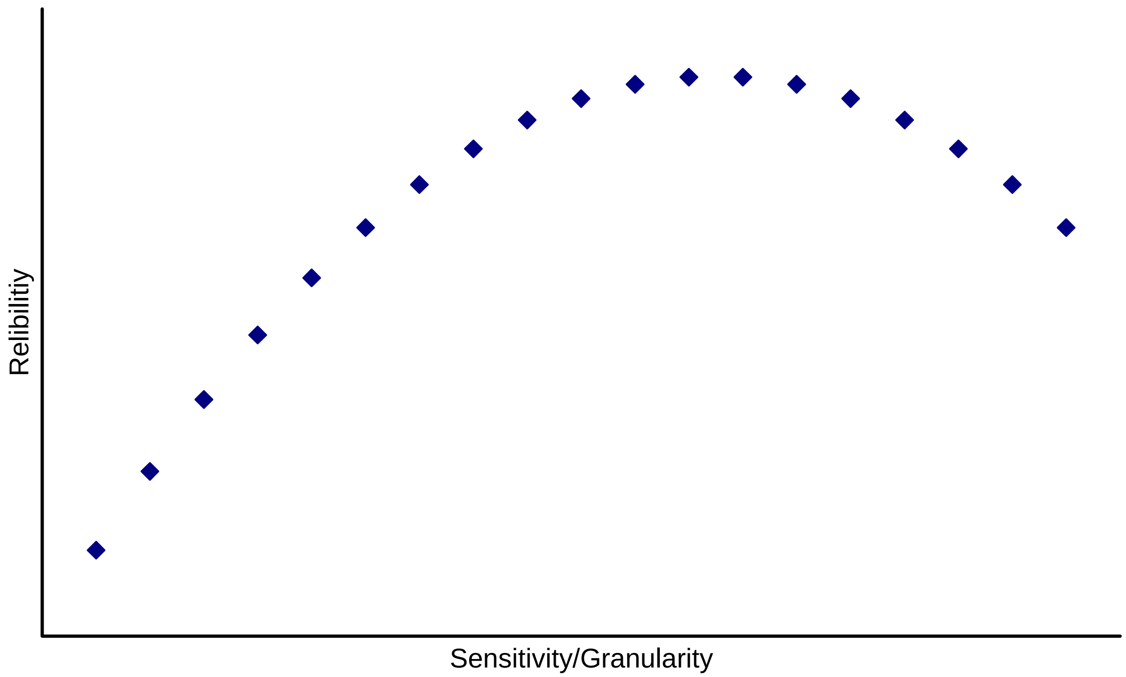
- Some Y's do not vary freely (due to conventions, institutions)

  - Ex: President's party; prime time news analysis
  - Measuring Y in nested (shorter) units adds no information
  - Measuring Y in longer units throws away information

- Other Y's can vary freely

  - Use theory and empirical tests to establish NTU
    - One possible test would compare a candidate unit (call this U_{cand}) to a shorter unit (call this U_{short}). One would hope to find no systematic variation in U_{short} across the U_{cand} units (i.e., the intraclass correlation, rho, would be 0). If one had data on Xs gathered at the U_{short} level as well as U_{short} –level data on Y, one would also hope to find that none of the Xs can explain variation in Y once fixed effects for U_{cand} are specified.
  - Tradeoff sensitivity/reliability: select the shortest unit that satisfies the criterion above up to the point that doing so begins to entail reliability costs

Reliability of Measure (Y axis) by Sensitivity of Measure (X axis)



Number of temporal units in a year-long long research period:

Months	n=12
Weeks	n=52
Days	n=365
Hours	n=8760
Minutes	n=525,600
Seconds	n=31,536,000
Shakes[1]	n=3,153,599,999,996,478

[1] A shake is 10 nanoseconds, roughly equivalent to the lifespan of a neutron. Researchers have only recently developed the technology to measure atomic behaviour across such tiny temporal units. See http://www.smartplanet.com/blog/smart-takes/ibm-claims-nanotech-breakthrough-atoms-measured-in-nanoseconds/10905 (accessed 8/8/11).

- In Ys that vary freely, selection of a longer Y-unit than NTU will sometimes (often?) be necessary because of feasibility constraints

- Example 1:

  - Y: Extent of President's speech on terrorism
  - Xs: Terrorist incidents (daily); Lagged presidential approval (monthly); Party of President (quadriennially)

  NTU of Y: day—feasible

- Example 2:

  - Y: Opinion on Social Security, varying across people and time
  - Hypothesis: Opinion shifts abruptly at age 65
  - Sample: Individuals aged 63-65
  - Research period: 1 year. 1/3 of individuals will remain <65, 1/3 will turn 65 (on different days), 1/3 will be >65 to start.

  NTU of Y: day—infeasible (sample every Kth day)

- “Natural interval” of a flow
- Sometimes flows can be derived from stocks
  - T is stock-NTU; interval from T to T+1 is flow-unit
- Not so in research on acts or events (e.g. riots, wars)
  - These are flows because they emerge between two time points
  - “+1” (“duration” of event) not obvious
  - NTU= unit of “minimum possible event duration?” Often not a viable criterion (e.g. acts of protest), to put it mildly. Intervals too short. Impractical for data collection. Many zeros (no events, no variation at all). Nothing of theoretical interest to explain across such tiny intervals. (What explains variation in time of day in which protest occurs? Do we care?)

- Solutions:

1. Move up to a longer and practicable interval if no signification information loss (most variation between and not within units)

   – Example: protest acts measured daily even if acts may last less
     - Very few cases of protest lasting more than a day
     - Variation in “time of the day” unlikely to be causally meaningful

2. If data constraints, aggregate across a longer interval but

   – explain what the natural interval is and why it is not being used,
   – be more modest or cautious in the conclusions drawn

3. Avoid temporal units altogether

   – When studying event characteristics (e.g., duration of or damage wrought by a riot), use events as units of analysis

   – Consider event sequence analysis, where actions or events sequenced in time (e.g., decision 1, decision 2, decision 3) serve as the units of analysis.

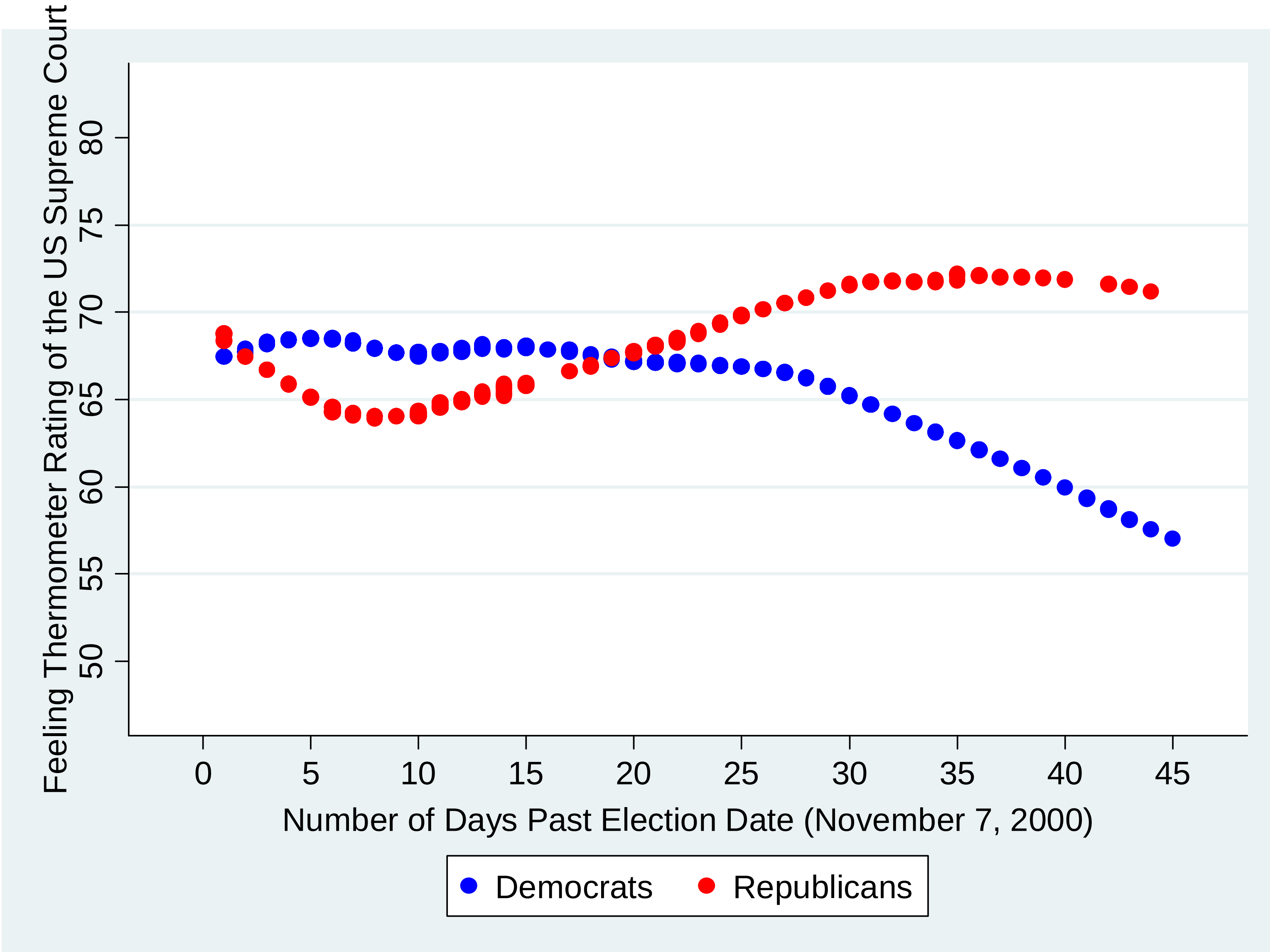
III. Missed Opportunities for Temporal Disaggregation

I. Longitudinal Variation in Cross-Sectional Surveys
II. Temporal Disaggregation in Difference in Difference Analyses

***

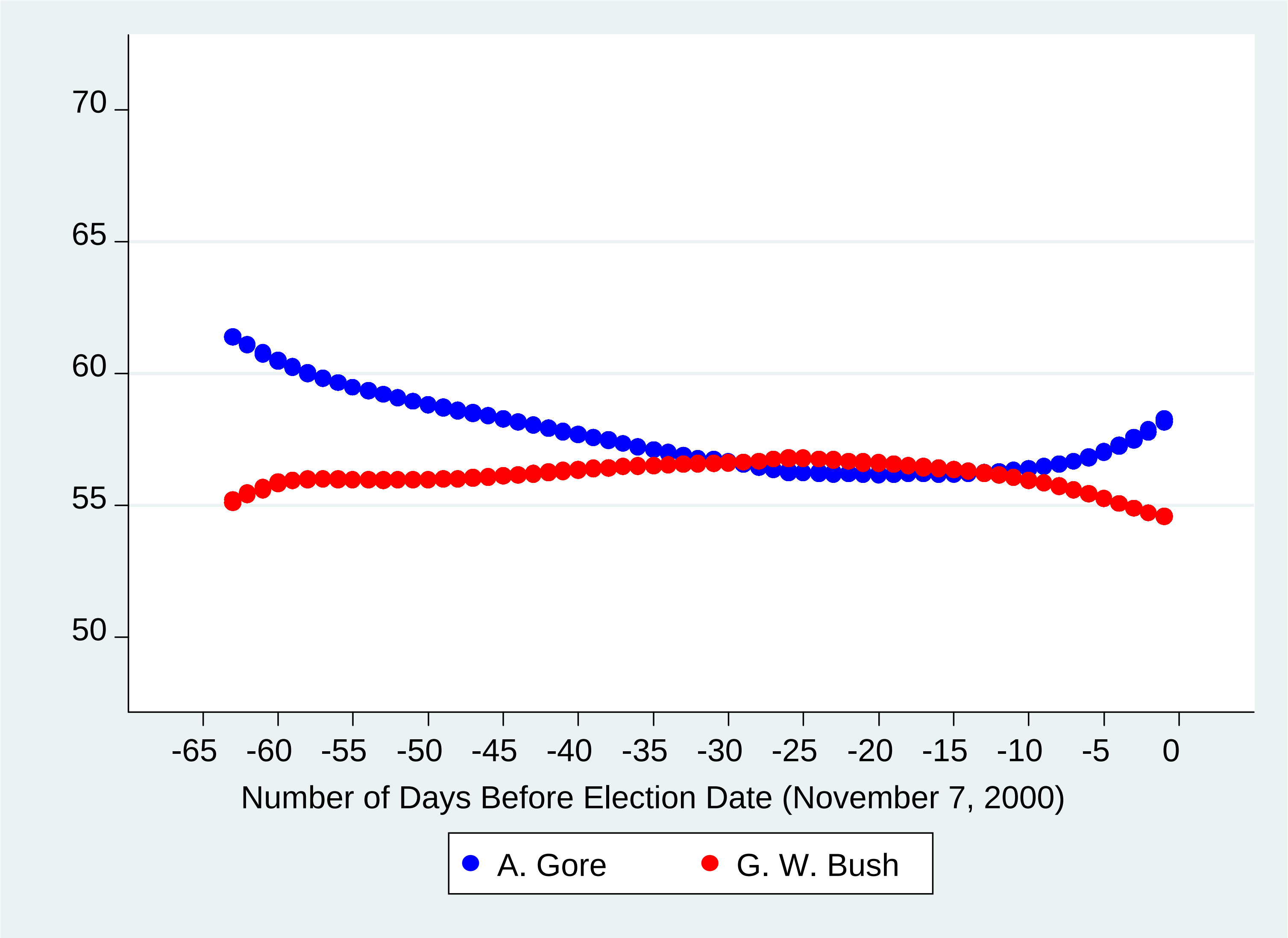
I. Longitudinal Variation in Cross-Sectional Surveys

Long field periods and randomized release of sample

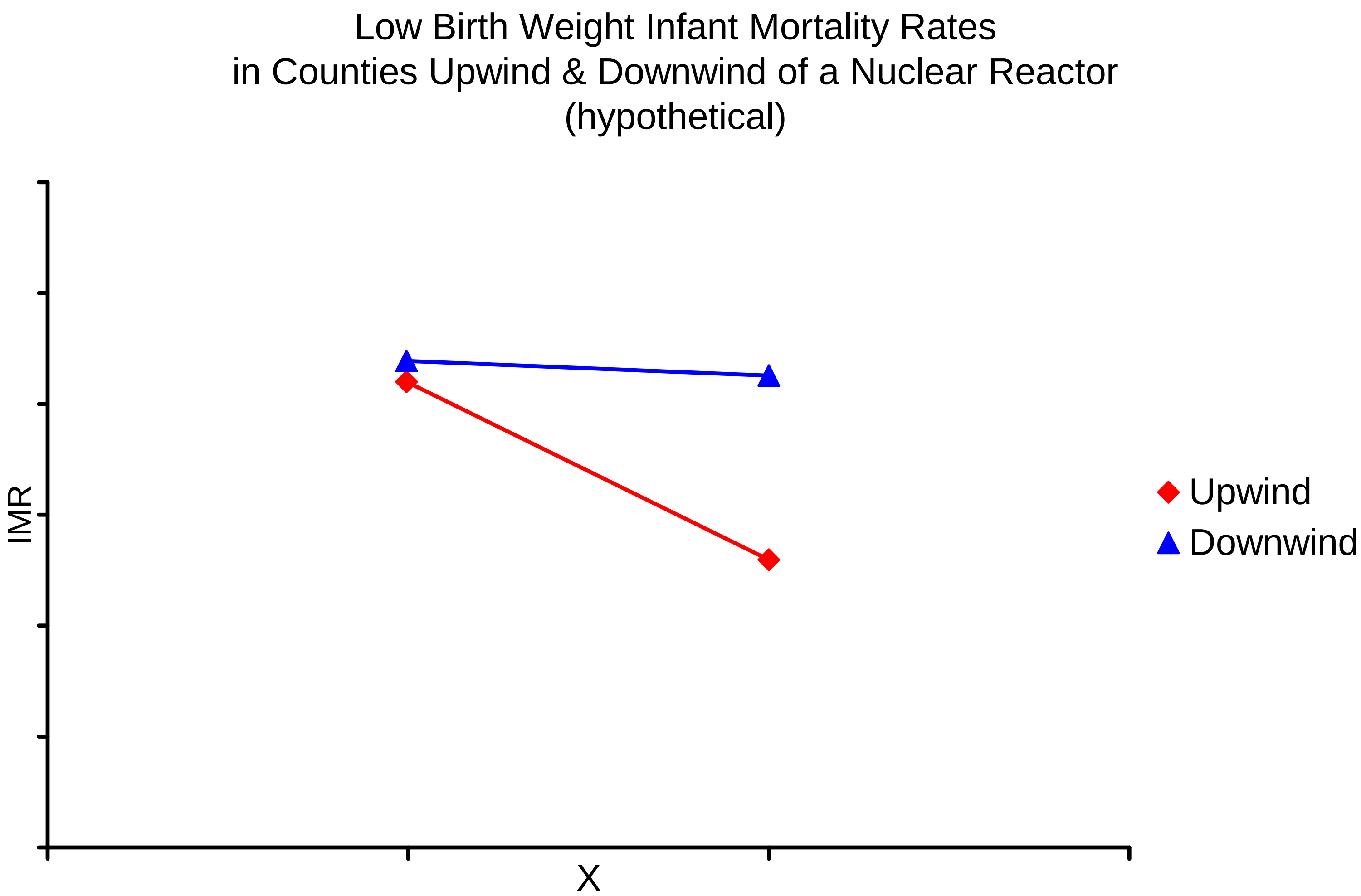


Feeling Thermometer Rating of the US Supreme Court
50
55
60
65
70
75
80
0
5
10
15
20
25
30
35
40
45
Number of Days Past Election Date (November 7, 2000)
Democrats
Republicans

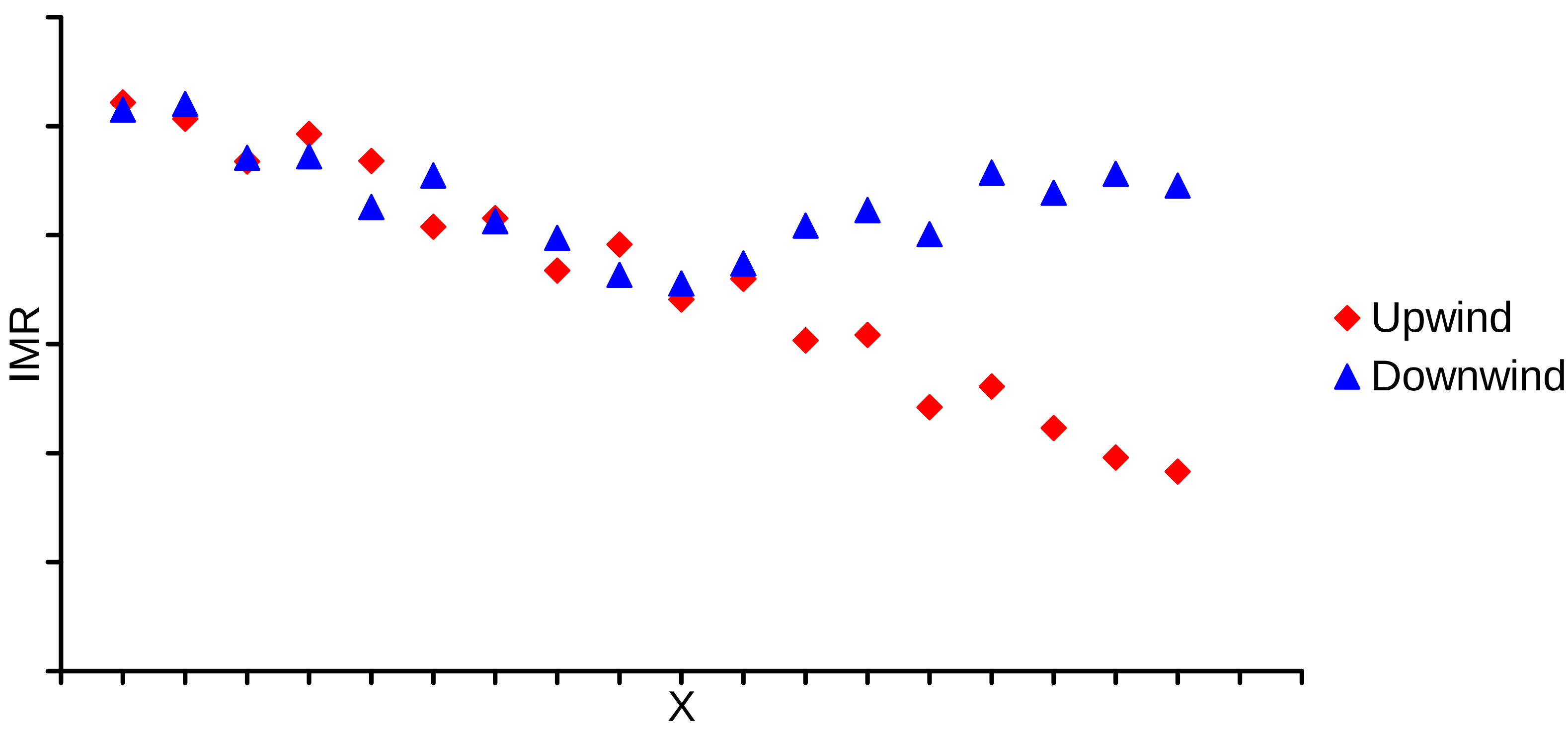
Pre-Election Ratings of the 2000 Presidential Candidates

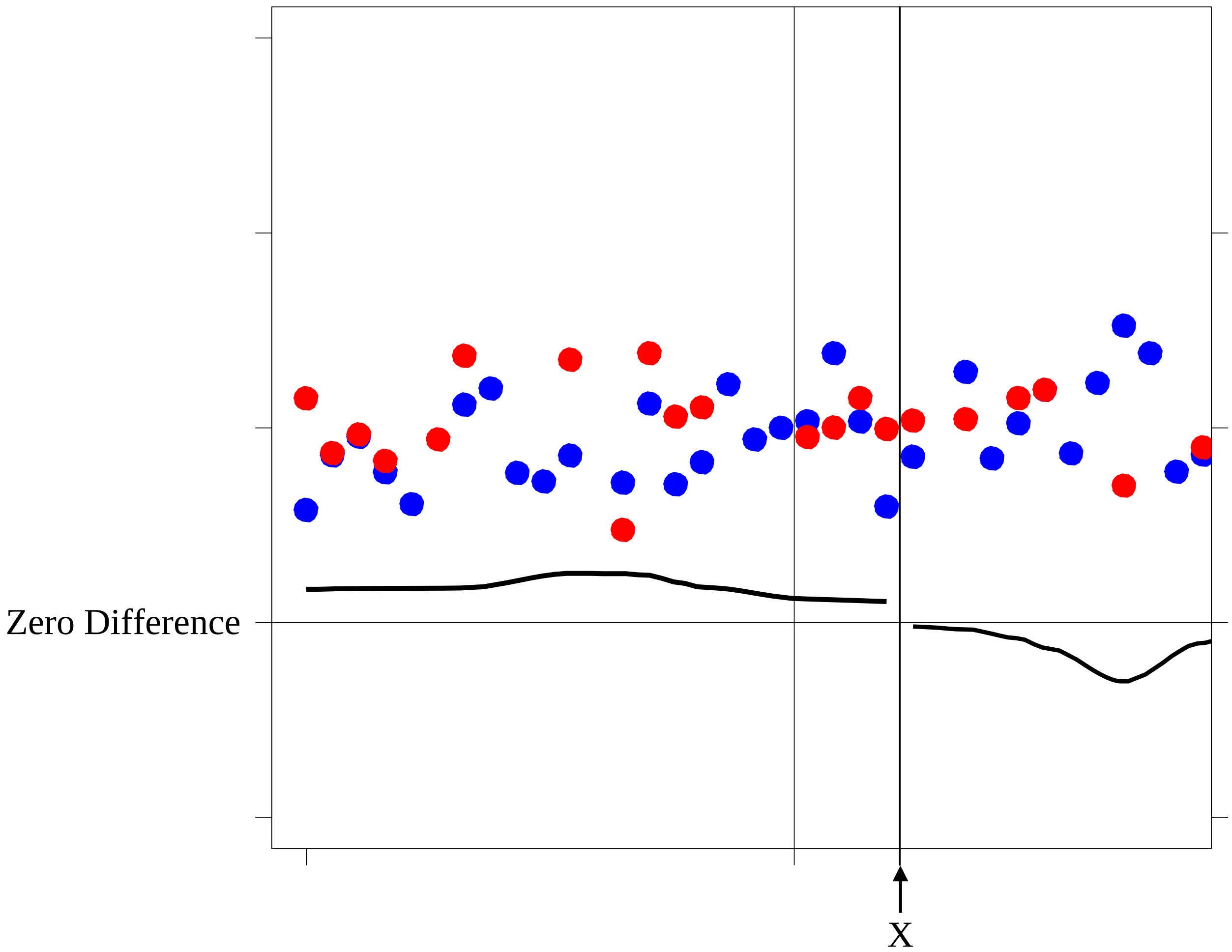


Temporal Disaggregation in Difference in Difference Analyses

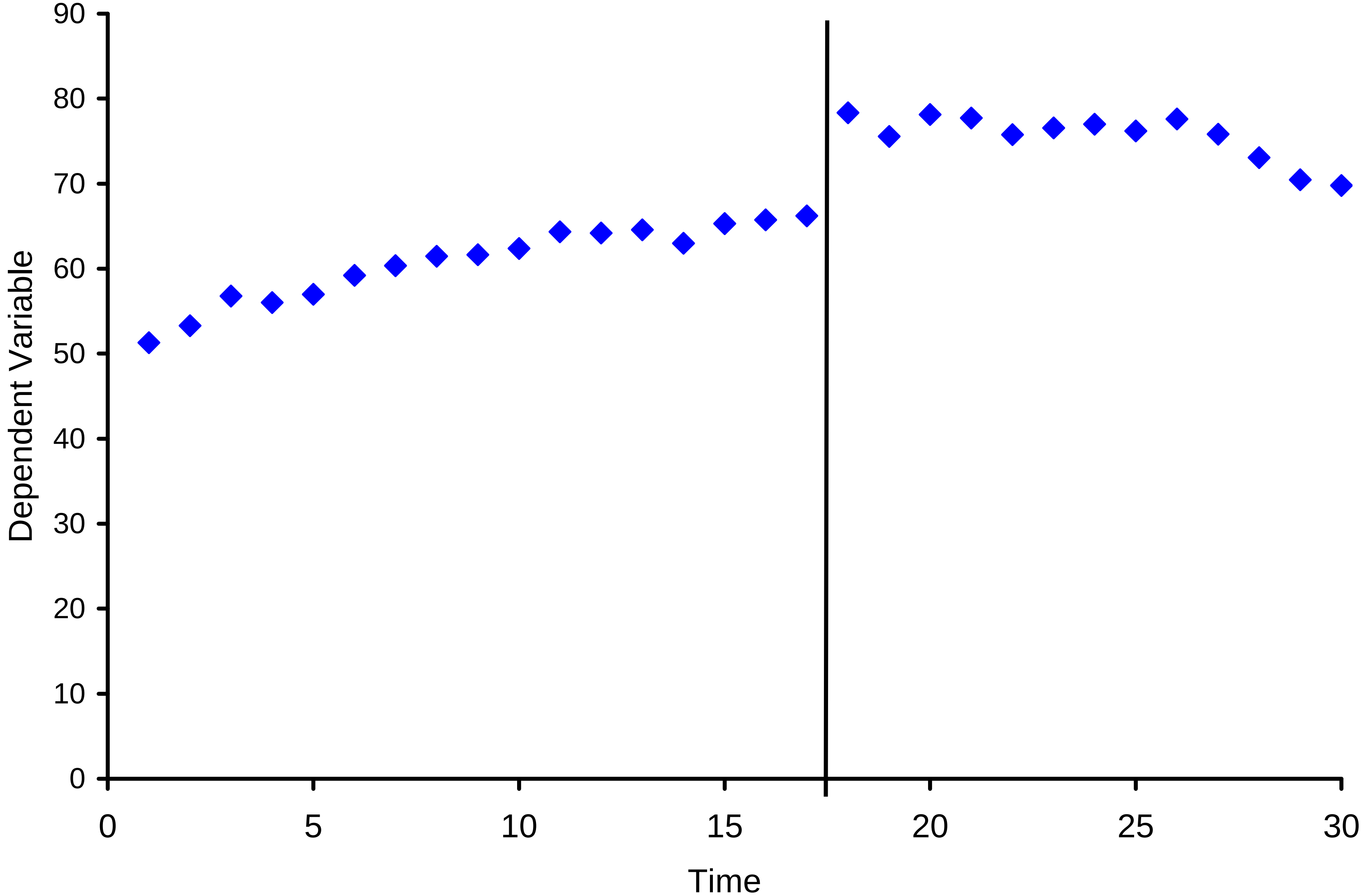


Low Birth Weight Infant Mortality Rates
in Counties Upwind & Downwind of a Nuclear Reactor
(hypothetical)

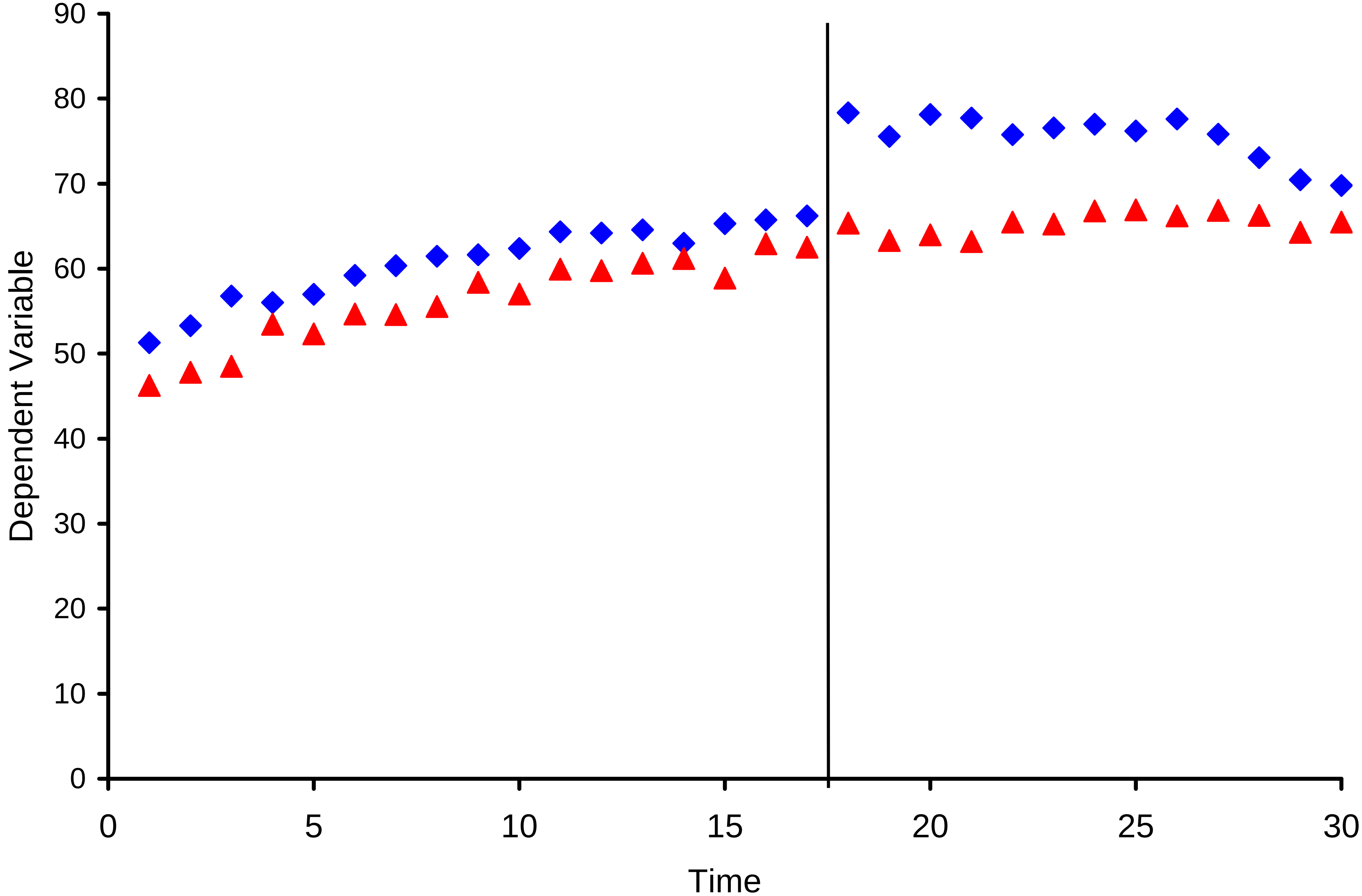




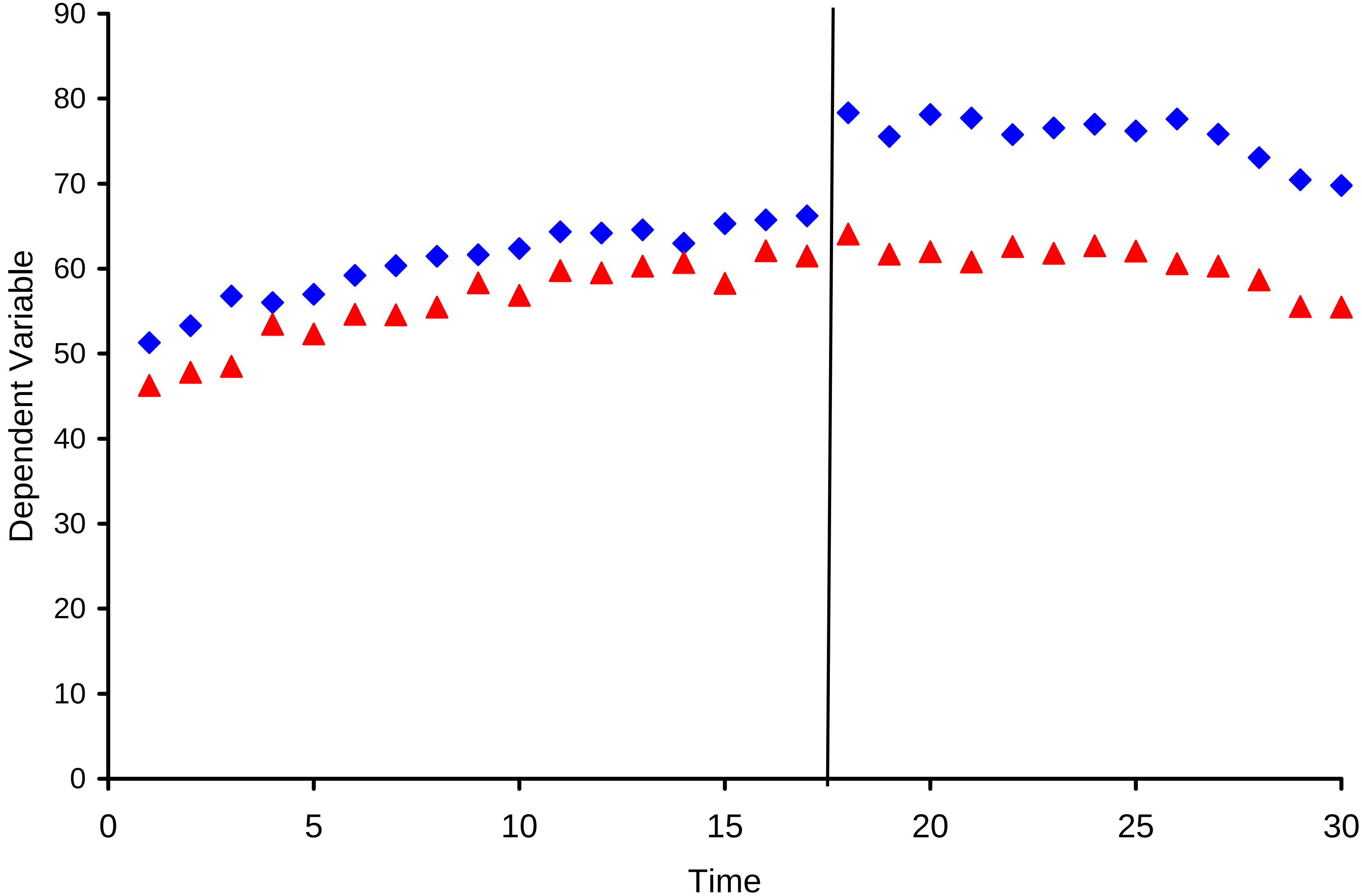
Zero Difference
X



90
80
70
60
50
40
30
20
10
0
Dependent Variable
0
5
10
15
20
25
30
Time



90
80
70
60
50
40
30
20
10
0
Dependent Variable
0
5
10
15
20
25
30
Time



Dependent Variable
Time
0
10
20
30
40
50
60
70
80
90
0
5
10
15
20
25
30

Thank you!