1 Course Outline

- Course Requirements
- Office Hours
- Syllabus

Mostly Empirical Political Economy

Fall, 2009
Columbia University
Committee on Global Thought, Economics, Political Science

This is a second year Ph.D. course in Political Economy. The purpose of the course is to introduce doctoral students to the field of political economy while at the same time introducing students to a wide range of empirical methods.

Before there was "economics" and "political science", there was political economy. In the past 20-30 years, there has been a substantial literature in what is now called positive political economy. Until 10 years ago, this literature was primarily theoretical. However, empirical political economy has been a very active field of research in the past 10 years.

There are two ways in which 'political economy' is used. The first is as the study of interactions between the economy and the political system. The second use of political economy is the use of economics methods (models and econometrics) to answer political questions.

Political Economy is divided into two distinct areas:

(1.) Theoretical Work – Models, usually of individual actors interacting with economic and political institutions
(2.) Empirical Work – Applied Econometrics

The current course will consider the second of these. However, other courses taught in the spring will cover the first. In the field of political science, political economy models are usually referred to as rational choice models.

In terms of empirical work, there are many different approaches. The first approach is the approach that this course will mostly focus upon. This is the natural experiment approach where existing data is used to identify causation. The second approach is the experimental approach. In this approach, data is created in order to identify causation. There are two types of experiments: field and laboratory. I will cover the former. However, Alessandra Cassella will cover the latter in the spring. Lastly, there is the structural approach where data is used to estimate parameters of models. There is less focus on causation. It is my view that empirical should always be concerned with causal identification. However, most questions are difficult to causally answer in a satisfactory way. I do believe that there is a tradeoff between importance of a topic and ability to identify causation. The course will therefore discuss a wide range of papers, some of which are very convincing in terms of their identification (i.e. the impact of different strategies to turn out voters) and others which attempt seriously, though with less success, to address causation on very difficult and important topics (i.e. the impact of institutions on growth).

The course will cover many topics:

We begin with political preferences. Preferences interact with economic and political institutions, leading to social outcomes. We will discuss what constitutes an appropriate model for preferences and how preferences are formed. This includes preferences over whether or not to vote, preferences over who to vote for (partisanship), and preferences over ideology. While dealing with these topics, we will introduce the experimental method, ordinary least squares estimation (OLS), and matching estimators.

We then will turn to politician preferences. A large theoretical literature claims (the Downsian competition literature) that policies are determined by voters, not politicians. However, there is a large body of recent work that shows that candidate gender, race, and political views all may affect the policies they support as well as the policies that are implemented by the political system. In this section of the course, we will introduce event study analysis and the regression discontinuity estimator.

Having looked at voter preferences and politician preferences, we will then turn to a third important force on political outcomes: the impact of money in politics (i.e. special interest politics). Here, we will introduce fixed effects estimation.

One way for money to influence politics is by disseminating information. However, a theoretical literature claims that in the long run, rational (in the sense of rational expectations) decision makers should not be influenced by a biased source of information. We will discuss the impact of the media on preferences. There is a large debate on whether the media is demand or supply driven. In this part of the course, we will introduce random effect estimators, instrumental variables, a generalization of the IV estimator called the control function approach, and non-independence of errors.

We then will look at how incentives within the political system impact the performance of politicians. As part of this, we will look at how politicians may use debt to influence future politicians. We will also discuss stylized facts about macro political economy (the so-called political business cycle).

Then we will begin our study of the impact of political institutions. We begin with a discussion of individual versus institutions. We look at the long run impacts of individual leaders on the degree of democracy, war, and economic growth.

We then discuss the institutional impacts of voting rules. In particular, we will look at majoritarian versus proportional systems, representative democracy and the secret ballot.

We then turn to the political determinants of growth. In particular, we consider the role of slavery, colonialism, and economic property rights on growth. There is a large literature in both political science and economics on the impact of democracy on growth, with some saying that growth leads to democracy and others saying that democracy leads to growth. We will also look at this literature.

One of the possible reasons for slow growth is violence. A large literature both in political science and in economics looks at the causes of violence. We will focus on the economic determinants of civil conflict. We will also discuss weak instrumental variable problems.

Lastly, we look at international relations. We will consider the impact of being a member of the UN security council on bilateral aid.

If there is time, we will also look at alternative distributional assumptions on functional form.

Course Requirements:

There will be two empirical problem sets and a original final research paper on empirical political economy. The purpose of the research paper is for you to work on something which could end up being a part of your dissertation.

Note: Two classes during the semester will have to be rescheduled because I will be away.

Papers

(papers with two ** will be read in depth, papers with one * will be discussed in less depth, and papers with no stars are additional relevant literature)

I. Preferences (Lectures I-II)

** Gerber, Alan and Donald Green (September, 2000), "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment", *American Political Science Review* Vol. 94, Num. 3, pp. 653-663.

** Imai, Kosuke (May, 2005), "Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments", *American Political Science Review* Vol. 99, Num. 2, pp. 283-300.

** Gerber, Alan, Donald Green and Christopher Larimer (2008), "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment", *American Political Science Review*, Vol. 102 (1), pp. 33-48.

* Duflo, Esther, Rachel Grennerster, and Michael Kremer, "Using Randomization in Development Economics Research: A Toolkit", working paper, http://econ-www.mit.edu/files/806

New Statistical Tools: Randomized Experiments
Matching

II. Preferences: Ideology and Partisanship (Lecture II)

* Alesina, Alberto and Nicola Fuchs-Schundeln (2007), “Good Bye Lenin (Or Not?): The Effect of Communism on People's Preferences”, *American Economic Review*, 97(4).

* Mullainathan, Sendhil and Ebonya Washington (2009), "Sticking with Your Vote: Cognitive Dissonance and Political Attitudes," *American Economic Journal: Applied Economics,* Vol. 1(1), pp. 86-111.

* Greg Duncan, Johanne Boisjoly, Michael Kremer, Dan Levy, and Jacque Eccles (2006), "Empathy or Antipathy? The Consequences of Racially and Socially Diverse Peers on Attitudes and Behaviors,", *American Economic Review,* Vol. 96(5), pp. 1890-1906.

* Cliningsmith, David, Asim Khwaja and Michael Kremer (2008), “Impact of the Hajj”, working paper.

New Statistical Tools: OLS/Natural Experiments

III. Preferences of politicians (Lecture III)

* Chattopadhyay, Raghabendra and Esther Duflo (2004), “Women as Policy Makers: Evidence from a Randomized Policy Experiment in India,”, *Econometrica* Vol. 72(5): 1409-1443.

* Washington, Ebonya (2006), "How Black Candidates Affect Voter Turnout," *Quarterly Journal of Economics*, 2006, pp. 121 (3).

** Fisman, Raymond (2001), "Estimating the Value of Political Connections." *American Economic Review* Vol. 91 (4), pp. 1095-1102.

Pande, Rohini, “Can Mandated Political Representation Provide Disadvantaged Minorities Policy Influence? Theory and Evidence from India,” *American Economic Review*, September 2003, Vol. 93(4): pp. 1132-1151.

Edlund, Lena and Rohini Pande (August, 2002), “Why Have Women Become Left-Wing? The Political Gender Gap and the Decline in Marriage,” with L. Edlund, *Quarterly Journal of Economics*, Vol. 117: 917-961.

Washington, Ebonya (2008), "Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues," *American Economic Review*, 2008, Vol. 98(1), 311-332.

New Statistical Tools: Event Study Methodology

IV. Partisanship and Policy Outcomes (Lectures IV-V)

** Lee, David S., Enrico Moretti and Matthew Butler (2004), "Do Voters Affect or Elect Policies? Evidence from the U.S. House ", Vol. 119(3).

* Lidbom, Per Petterson (2008), "Do Parties Matter for Economic Outcomes: A Regression-Discontinuity Approach," Journal of the European Economic Association, Volume 6, Issue 5, 1037–1056, 2008.

* Gyourko, Joseph and Fernando Ferreira (2009), "Do Political Parties Matter? Evidence from Cities", *Quarterly Journal of Economics,* Vol. 124 (1), pp. 349:397.

* Imbens, Guido and Thomas Lemieux (February, 2008), "Regression Discontinuity Designs: A Guide to Practice", *Journal of Economicetrcs*, Vol. 142(2), pp. 615-635.

New Statistical Tools: Regression Discontinuity

V. Money and Political Influence (Lecture V)

** Levitt, Steven (1994), "Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House", *Journal of Political Economy* Vol. 102, Num. 4, pp. 777-797.

New Statistical Tools: Fixed Effects

VI. Media, Information and Ideology/Partisanship (Lectures VI-VII)

** Gentzkow, Matthew and Jesse Shapiro (2006), *"What Drives Media Slant? Evidence from U.S. Daily Newspapers"*, working paper.

** Eisensee, Thomas and David Stromberg (May, 2007), "News Floods, News Droughts, and U.S. Disaster Relief", *Quarterly Journal of Economics*, 122(2), 2007. http://www.iies.su.se/~stromber/Disasters.pdf (use the working paper version).

* Snyder, James and David Stromberg (2008), "Press Coverage and Accountability", working paper.

* Bjorkman, Martina and Jakob Svensson (forthcoming), "Power to the People: Evidence from a Randomized Field Experiment of Community-Based Monitoring in Uganda", *Quarterly Journal of Economics*.

Stromberg, David (2004), "Radio's Impact on Public Spending", *Quarterly Journal of Economics*, Vol. 119(1), 2004.

New Statistical Techniques: Random Effects
Instrumental Variables
Control Function
Clustered Errors

VII. Politician Incentives (Lecture VII)

** Besley, Timothy and Anne Case, "Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits", Vol. 110, Num. 3, pp. 769-798.

** "An Empirical Investigation of the Strategic Use of Debt" (2001), *Journal of Political Economy*, 109, pp. 570-84.

VIII. Individuals Versus Institutions (Lecture VIII)

** Jones, Ben and Ben Olken, "Hit or Miss? The Effect of Assassinations on Institutions and War", working paper. http://www.nber.org/~bolken/assassinations.pdf

* Jones, Ben and Ben Olken, "Do Leaders Matter? National Leadership and Growth since World War II" Vol. 120(3) pp. 835-864.

IX: Political Institutions: Forms of Government and Voting Rules (Lecture IX)

** Pettersson-Lidbom, Per and Björn Tyrefors, "The Policy Consequences of Direct versus Representative Democracy: A Regression Discontinuity Approach", working paper.
http://people.su.se/~pepet/directdem.pdf

** Aghion, Philippe, Alberto Alesina and Francesco Trebbi (2008), Electoral Rules and Minority Representation in US Cities, *Quarterly Journal of Economics*, February, Vol. 123(1): pp.325-357.

** Baland, Jean-Marie and James A. Robinson (2008), “Land and Power: Theory and Evidence from Chile," *American Economic Association*, pp. 1737-65.

X. Institutions and Growth (Lectures X-XI)

Growth and Development

** Feyrer, James and Bruce Sacerdote, “Colonialism and Modern Income: Islands as Natural Experiments”, working paper.
http://www.dartmouth.edu/~jfeyrer/islands.pdf

** Banerjee, Abhijit, and Lakshmi Iyer. "History, Institutions and Economic Performance: the Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95, no. 4 (September 2005): 1190-1213.

** Acemoglu, Daron, Simon Johnson and Jim Robinson (December, 2001), “The Colonial Origins of Comparative Development: An Empirical Investigation”, *American Economic Review*, Vol. 91, pp. 1369-1401.

* Albouy, David, “The Colonial Origins of Comparative Development: An Investigation of the Settler Mortality Data”, revised and resubmit at the *American Economic Review*.
http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1055&context=iber/cider

* Acemoglu, Daron, Simon Johnson and James A. Robinson, “Reply to the Revised (May, 2006) version of David “The Colonial Origins of Comparative Development: An Investigation of the Settler Mortality Data”, working paper.
http://econ-www.mit.edu/files/212

* Besley, Timothy and Torsten Persson (forthcoming), “The origins of state capacity: Property rights, taxation, and policy", *American Economic Review*.

* Nunn, Nathan (2008), "The Long Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, Vol. 123(1), pp. 139-176.

Alesina, Alberto, William Easterly and Janina Matuszeski (June, 2006), “Artificial States”, working paper.
http://www.nyu.edu/fas/institute/dri/Easterly/File/artificialstatesNBER.pdf

Iyer, Lakshmi. "Direct versus Indirect Colonial Rule in India: Long-term Consequences." *The Review of Economics and Statistics* (forthcoming).

Tilly, Charles (2007), **Coercion, Capital and European States: AD 990-1992**, Wiley-Blackwell.

XI. Democracy and Growth (Lecture XII)

** Daron Acemoglu, Simon Johnson, James A. Robinson and Pierre Yared (2008), "Income and Democracy", *American Economic Review*, 98(3), pp. 808-42.

* Besley, Timothy, Torsten Persson and Daniel Sturm (2006), "Political Competition, Policy and Growth: Theory and Evidence from the United States", working paper.

* Rodrik, Dani (August, 1999), "Democracies Pay Higher Wages", *Quarterly Journal of Economics*, Vol. 94, Num. 3, pp.707-738.

XII. Resource Curse and Violence (Lecture XIII)

** Miguel, E., S. Satyanath and E. Sergenti (2004), "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," Journal of Political Economy, 112(4), 725-753.

** Dube, Oeindrila and Juan Vargas (2008), "Commodity Price Shocks and Civil Conflict: Evidence from Columbia", working paper.

** Guidolin, Massimo and Eliana La Ferrara, "Diamonds Are Forever, Wars Are Not. Is Conflict Bad for Private Firms?", forthcoming *American Economic Review*.
http://www.igier.uni-bocconi.it/whos.php?vedi=1189&tbn=albero&id_folder=177

New Statistical Tools: Weak Instruments

XIII. International Politics & International Relations (Lecture XIV)

** Kuziemko, Ilyana, and Eric D. Werker (2006), "How Much Is a Seat on the Security Council Worth? Foreign Aid and Bribery at the United Nations", *Journal of Political Economy* Vol. 114(5), pp. 905-930.

Additional Methodological Paper:

Angrist, Joshua and Alan Krueger (Fall, 2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", *Journal of Economic Perspectives V.15 # 4,* pp. 69-85.

2 Identification of Causation

- Godel's Theorem : Any logical system capable of whole number addition and multiplication will have question which can be asked and can not be answered true or false. Moreover, it is impossible to determine which questions can not be proven true or false.

  - Empirical Analogue: Empirical questions can be asked which may likely have no good answer.

- Ability to Understand the Past and Predict the Future

  - Making Progress on Identification to Important Questions: Acemoglu, Johnson and Robinson

  - "Clean Identification": Lee, Butler, and Moretti

  - Establishing Empirical Facts: Shleifer et al., Persson and Tabellini, Barro

- Hume's Problem
  - Need for the Ceterus Paribus Assumption for Internal Identification
  - Need for the Ceterus Paribus Assumption for External Validity

3 Empirical Approaches

- Identification of Causality:Experiments
  - Field
  - Laboratory
- Natural Experiments
  - Randomization
  - Conditional Randomization
- Identification of Parameters: Structural Estimation
- Stylized Fact Provision: Partial Sample Correlation

4 Rubin's Potential Outcome Model (1974)

- Treatment T is binary: $\{0, 1\}$.

- Outcome for Treatmeant is given by random variable Y_i^T and outcome for control: Y_i^C

- Impact of Treatment given by Potential Outcomes: $TY_i^T + (1 - T)\, Y_i^C$

- Conditioning Sets (types who would select into control and treatment): $\{C, T\}$

- Treatment Effect Without Randomization:

$$E\left[Y_i^T - Y_i^C\right] = E\left(Y_i^T|T\right) - E\left(Y_i^C|C\right) =$$

$$\left\{E\left(Y_i^T|T\right) - E\left(Y_i^C|T\right)\right\} + \left\{E\left(Y_i^C|T\right) - E\left(Y_i^C|C\right)\right\}$$

{Treatment Effect (on the treated)} + {Selection Bias}

- Treatment Effect With Perfect Randomization (S is the Selection Criterion for the experiment):

$$E\left[Y_i^T - Y_i^C\right] = E\left(Y_i^T|S\right) - E\left(Y_i^C|S\right)$$

5 Experiments: An Introduction

- Experiments: Two Types
  - Laboratory: i.e. Iyengar, Going Negative, showing face pictures of politicians
  - Field: i.e. Bjorkman and Svensson: Randomly providing and making public report cards about health clinic performance

- Benefits
  - Ability to control for selection
  - Ability to design an experiment to ask exactly the question you wanted to answer
  - Ability to commit to a research design ahead of time and reduce degrees of freedom for manipulation

        * Subgroups (variables and/or strata) as Degrees of Potential Manipulation

        * Tradeoff Between Ex-Post Learning and Ex-Ante

- Costs

    - External Validity

        * Moral Constraints

        * Legal Constraints

            · Attrition

            · Substitution Bias (Heckman and Smith, 1995)

            · Randomization Bias and Selection in Experiment Participation (Heckman and Smith, 1995)

        * Effects of the Experiment Independent of the Treatment

- Hawthorne Effects: Changes in behavior among the treatment group

- John Henry Effects: Changes in behavior among the control group

– Internal Validity

∗ Attrition

∗ Externalities: SUTVA (Stable Unit Treatment Value Assumption) "General Equilibrium": $E(Y_i^k|T)$ is independent of T_j.

$$\begin{aligned} E\left(Y_i^T|T\right) - E\left(Y_i^C|C\right) &= \\ E\left(Y_i^T - Y_i^C|T\right) &= E\left(Y_i^T - Y_i^C\right) \end{aligned}$$

∗ Contamination

- Treatment Doesn't Take Up

- Control Takes Up

– Small Sample Sizes:

    * Power
    * Identifying Heterogeneous Effects
    * Identifying Population Average Treatment Effects
  - Monetary Costs of Implementation

6 Power Calculations

$$Y_i = \alpha + \beta T_i + \epsilon_i$$

OLS Estimator is:

$$\min_{\alpha,\beta} \sum_{i=1}^{I} (Y_i - \alpha - \beta T_i)^2$$

$$\begin{aligned}
\alpha &: -2 \sum_{i=1}^{I} (Y_i - \alpha - \beta T_i) = 0 \\
\beta &: -2 \sum_{i=1}^{I} (Y_i - \alpha - \beta T_i) T_i = 0
\end{aligned}$$

This implies the following estimators for α and β :

$$\begin{aligned}
\alpha &= \frac{\sum_{i=1}^{I} (Y_i - \beta T_i)}{I} = \bar{Y} - \beta \bar{T} \\
\hat{\beta} &= \frac{\sum_{i=1}^{I} \left(Y_i - \bar{Y}\right) \left(T_i - \bar{T}\right)}{\sum_{i=1}^{I} \left(T_i - \bar{T}\right) \left(T_i - \bar{T}\right)}
\end{aligned}$$

Generally (with spherical disturbances): Standard Errors Given By:

$$\sigma_{\epsilon}^{2}\left(X'X\right)^{-1}$$

We can also compute the Standard Error by taking $V\left(\hat{\beta}\right)$ (and remembering that $P\left(T=1\right)=P$):

$$\frac{\sigma_{\epsilon}^{2}}{P\left(1-P\right)N}$$

- Size and Power
  - Size of a Test: Probability of a Type I Error (Probability of Failing to Reject a True Null) = 1 - Confidence Level.
  - Power of a Test: 1 - Probability of a Type II Error

(Probability of Rejecting a False Null Hypothesis)

	Do Not Reject H_0	Reject H_0
H_0 is True	Correct Decision $1-\alpha$: Confidence Level	Type I Error Size of Test α : Significance Level
H_0 is False	Type II Error ω	Correct Decision $1-\omega$: Power of Test

- In order to reject a null hypothesis of no effect at an α level of confidence, we need:

$$\hat{\beta} > t_\alpha SE\left(\hat{\beta}\right)$$

- If we want power of $1-\omega$:

$$\hat{\beta} > \left(t_{1-\omega} + t_\alpha\right) SE\left(\hat{\beta}\right)$$

- $SE\left(\hat{\beta}\right) = \sqrt{V\left(\hat{\beta}\right)} = \sqrt{\frac{\sigma_\epsilon^2}{P(1-P)N}}$

- Therefore the Minimum Detectable Effect (MDE) where α is the size and $1-\omega$ is the power:

$$\hat{\beta} > (t_{1-\omega} + t_{\alpha})\sqrt{\frac{\sigma_{\epsilon}^{2}}{P(1-P)N}}$$

- Now suppose we have grouped data with group effects: v_j. Then, we estimate:

$$Y_{ij} = \alpha + \beta T_{ij} + v_j + \epsilon_{ij}$$

  – where there are J clusters of size n

  – $v_j\tilde{\ }i.i.d.\ N\left(0,\tau^2\right)$ and $\epsilon_{ij}\tilde{\ }i.i.d.\ N\left(0,\sigma_{\epsilon}^2\right)$

  – Then we get as our MDE:

$$\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{n\tau^2+\sigma_{\epsilon}^2}{nJ}}$$

  – With individual randomization, we would get:

$$\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\tau^2+\sigma_{\epsilon}^2}{nJ}}$$

- Ratio between the two $= D = \sqrt{1 + (n-1)\rho}$ where $\rho = \frac{\tau^2}{\tau^2 + \sigma_\epsilon^2}$

- With imperfect compliance:

$$\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma_\epsilon^2}{N}}\frac{1}{c-s}$$

- $V(Uncond.) > V(Cond.) > V(Stratified)$: Imbens, King and Ridder.

- With stratification, this is a diagonal matrix in which case adding a dimension of stratification (constructed to be orthogonal to the other dimensions) will always reduce the standard errors.

- If we want power of $1 - \omega$:

7 Intention to Treat Estimates in Experiments

- Treatment: T, Assignment of Treatment: Z

- Average Treatment Effect (ATE): $E\left(Y_i^T - Y_i^C\right)$

  - If you can actually randomize treatment

- Intention To Treat (ITT): $E\left(Y_i^T - Y_i^T|Z\right) = E\left(Y_i^T|Z = 1\right) - E\left(Y_i^T|Z = 0\right)$

  - If you can randomize access to Treatment but not Treatment itself

- Local Average Treatment Effect (LATE): Will discuss later.

- Do we want the intention to treat estimate or the treatment effect?

1 Matching Estimators: Motivation

- What are matching estimators?
  - Individual Matching: Match observations and estimate between individually matched observations.

    $$\sum_{i=1}^{N} w_i \left[Y_i \left(T = 1 \right) - Y_{M(i)} \left(T = 0 \right) \right]$$

    where N is the number of treated observations, $Y_i \left(T = 1 \right)$ is the outcome for the i^{th} treated observation, $Y_{M(i)} \left(T = 0 \right)$ is the matched observation for the i^{th} treated observation, and w_i is the population weight of the i^{th} treated observation. (Note that each treated observation is matched to at most one untreated observation)

  - Block Matching: Match groups of similar observations (on covariates):

$$\sum_{i=1}^{N} w_i \left(\begin{array}{c} \frac{\sum_{j=1}^{Z_{i,T=1}} Y_{ij}(T=1)}{Z_{i,T=1}} - \\ \frac{\sum_{j=1}^{Z_{i,T=0}} Y_{M(i)}(T=0)}{Z_{i,T=0}} \end{array} \right)$$

where Y_{ij} is the j^{th} observation in the i^{ih} group and $Z_{i,T=1}$ is the number of observations in the i^{th} treatment group (and similarly for $Z_{i,T=0}$)

- Main questions with matching estimators:
  - How to do matching?
  - How to compute the weights (w_i)?
  - How to compute the standard errors?

- Many methods:
  - Criterion

    * Exact covariate matching

    * Propensity score matching

    * Mahalanobis matching

$$\min_{\{ik,jk\}} \sum_{\{ik,jk\}} \sum_{k} \left(X_{ik} - X_{jk}\right) S^{-1} \left(X_{ik} - X_{jk}\right)$$

  - Matching Techniques

    * Nearest neighbor matching (For every treatment, find the nearest control - with or without replacement)

    * Genetic algorithm matching

    * Many others!

- Why use matching estimators? Under Gauss Markov Assumptions, OLS is BLUE (Best Linear Unbiased Estimator)

- $Y_{it} = \alpha + \beta X_{it} + \epsilon_{it}$
- $cov\left(X_{it}, \epsilon_{it}\right) = 0$
- $cov\left(\epsilon_{it}, \epsilon_{jt}\right) = 0$ where $i \neq j$
- $V\left(\epsilon_{it}\right) = \sigma^2$

- Matching Estimators are linear in treatment (and thus in variables since matching estimators usually only include treatment variables as RHS variables). So why use matching estimators as opposed to OLS?

- Tradeoff: Efficiency vs. Robustness
  - What does robustness mean?
    - Must know ALL the relevant covariates
    - Lack of knowledge of functional form
    - Must have alot of data

    * Then conditional on covariates, only average difference in outcome comes from treatment. However, any specific functional form for OLS or NLS may be mis-specified. So, look only within covariate groupings.

- Problem: No models (that Im aware of) of robustness

- Other possibilities of more robust, local estimators with less variance (greater efficiency): non-parametric and semi-parametric matching

2 Matching Estimators: Estimation

- Use Rubin's Potential Outcomes Model

- Assume:

  1. Unconfoundedness: $(Y_i(1), Y_i(0)) \amalg W_i | X_i$
     - also called selection on observables assumption
     - twins example
  2. Overlap: $0 < \Pr(W_i = 1|X_i) < 1$
  3. (1) & (2) together are called "strongly ignorable treatment"

- Given strongly ignorable treatment:

$$E[Y_i(1) - Y_i(0)|X_i = x] =$$

$$E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x] =$$

$$E\left[Y_i(1)\,|X_i = x, W_i = 1\right] - E\left[Y_i(0)\,|X_i = x, W_i = 0\right] =$$

$$E\left[Y_i|X_i = x, W_i = 1\right] - E\left[Y_i|X_i = x, W_i = 0\right]$$

- How Different is Matching from OLS really?
  - Defintion: saturated models are models where there is a dummy variable for every covariate realization.
  - Example: LHS: Wages, RHS: Education (University Completion, HS Completion, Less Than High School), Race (Black, White), Sex (Female, Male). Transform variables into dummies (11 dummy variables with one category left out as the constant):
    1. University Completion, Black, Female
    2. University Completion, Black, Male
    3. University Completion, White, Female

4. University Completion, White, Male
5. HS Completion, Black, Female
6. HS Completion, Black, Male
7. HS Completion, White, Female
8. HS Completion, White, Male
9. $<$ HS, Black, Female
10. $<$ HS, Black, Male
11. $<$ HS, White, Female
12. $<$ HS, White, Male

– This replicates exact covariate matching though with different weights than matching estimators:

– Matching: (From Mostly Harmless Econometric, Angrist and Pischke):

$$\widehat{\beta}_T^{Match} = \frac{\sum_x E_x R_x}{\sum_x R_x}$$

where

$$\begin{aligned} R_x &= P(W_i = 1 | X_i = x) P(X_i = x) \\ E_x &= E[Y_{i1} | X_i, W_i = 1] - E[Y_{i1} | X_i, W_i = 0] \end{aligned}$$

– OLS:

$$\widehat{\beta}_T^{OLS} = \frac{\sum_x E_x R_x [1 - P(W_i = 1 | X_i = x)]}{\sum_x R_x [1 - P(W_i = 1 | X_i = x)]}$$

– So, OLS weights by variance of the observations, matching estimators by their population frequency.

– Is this a fair characterization of the differences between OLS and Matching? So, is matching a weighted OLS?

- Covariate Balance (Rosenbaum and Rubin, 1985)
  - With discrete variables and large samples, we may be able to look within actual covariate bins
  - With small sample or continuous variables, we will not be able to match exactly. But:
    * Matching should lead to covariate balance across treatment and control: $X_{i,T} - X_{M(i),T} \tilde{} N\left(0, \sigma^2\right)$
    * In other words, covariates should be randomly distributed across treatment and control for matched observations
- The Search Problem
  - Many search problems have an exponential asymptotic
  - Finding the optimal set of matches is such a problem.

- E.G., we want to estimate ATE and do matching without replacement:
- With 10 treated and 20 control obs: 184,756 possible matches
- With 20 treated and 40 control obs: 13,784,652,8820 possible matches
- With 40 treated and 80 control obs: 1.075e+23
- With 185 treated and 260 control: 1.633e+69 with 185 treated and 4000 control: computer infinity
- Matching with replacement makes the search problem explode even more quickly.

• Propensity Score Matching (Rosenbaum and Rubin, 1985) : Dimensional Reduction

– Suppose that unconfoundness holds, then:

$$(Y_i(1), Y_i(0)) \amalg W_i | p(X_i)$$

where $p(X_i)$ is the probability of treatment (estimated via linear probability, probit, or logit)

– Intuition: by leaving out covariates, we introduce ommitted variables bias. However, since $X_i \amalg W_i | p(X_i)$ (covariate balance), adding X_i will not change the estimate of Y_i on X_i

- Steps for propensity score estimation

1. First estimate selection equation

$$T_i = F(X_i) + \epsilon_i$$

2. Estimate fitted probabilites of selection

$$\hat{F}(X_i)$$

3. Create bins of a given width (or do nearest neighbor matching)

4. Check for covariate balance across treatment and control within bins

5. Estimate difference between treatment and control within bins

$$\frac{\sum_{j=1}^{M} Y_{ij}\,(T=1)}{Z_{i,T=1}} - \frac{\sum_{j=1}^{M} Y_{M(i)}\,(T=0)}{Z_{i,T=0}}$$

6. Choose weights

   (a) Homogeneous treatment effect: weight by size of bin or other measures of variance of estimate

   (b) Heterogeneous treatment effect: weight using population weights

7. Estimate average treatment effect by weighting across bins:

$$\sum_{i=1}^{N} w_i \left[\frac{\sum_{j=1}^{M} Y_{ij}\,(T=1)}{Z_{i,T=1}} - \frac{\sum_{j=1}^{M} Y_{M(i)}\,(T=0)}{Z_{i,T=0}}\right]$$

8. Estimate standard errors for average treatment effect

(a) Estimate component by component

$$\left[\text{from } \sigma^2_{X,T=1},\ \sigma^2_{X,T=0},\ \bar{Y}_{T=1},\ \bar{Y}_{T=0},\ p(x)\right]$$

(b) Bootstrap

- Problems with bootstrapping IV estimators (at least nearest neighbor matching) due to non-linearities: Abadie, Imbens (2006)

- Some problems with FE: Interlude on Measurement Error (From Ashenfelter and Krueger, 1994)
  - Attenuation:

    True:

$$y_i = \beta x_i + \epsilon_i$$

    Observed

$$\bar{x}_i = x_i + \delta_i$$

$$\hat{\beta} = \frac{\sum_{t=1}^{t} x_i y_i}{\sum_{t=1}^{t} x_i^2} =$$

$$\frac{\sum_{t=1}^{t} \bar{x}_i \left(\beta x_i + \epsilon_i\right)}{\sum_{t=1}^{t} \bar{x}_i^2} = \frac{\sum_{t=1}^{t} \left(\beta x_i^2 + \beta \delta_i x_i + x_i \epsilon_i + \delta_i \epsilon_i\right)}{\sum_{t=1}^{t} \left(x_i^2 + 2 x_i \delta_i + \delta_i^2\right)}$$

$$E\hat{\beta} = \frac{\beta \sum_{t=1}^{t} x_i^2}{\sum_{t=1}^{t} \left(x_i^2 + \delta_i^2\right)} = \frac{\beta \sigma_x^2}{\sigma_x^2 + \sigma_\delta^2} = \beta \left(1 - \frac{\sigma_\delta^2}{\sigma_x^2 + \sigma_\delta^2}\right) < \beta$$

- – $\frac{\sigma_\delta^2}{\sigma_x^2 + \sigma_\delta^2}$ is called the reliability ratio

– Measurement Error Tradeoff

Suppose T=2;

$$y_{1i} = \alpha Z_i + \beta X_{1i} + \mu_i + \epsilon_{1i}$$

$$y_{2i} = \alpha Z_i + \beta X_{2i} + \mu_i + \epsilon_{2i}$$

$$\mu_i = \gamma X_{1i} + \gamma X_{2i} + \delta Z_i + \omega_i$$

Then $\hat{\beta}_{FE} = \hat{\beta}_{FD}$ comes from the regression $y_{1i} - y_{2i} = \beta\left(X_{1i} - X_{2i}\right) + \epsilon_{1i} - \epsilon_{2i}$

It can be shown that

$$\hat{\beta}_{FE} = \beta\left(1 - \frac{\sigma_\delta^2}{\left[\sigma_x^2 + \sigma_\delta^2\right](1 - \rho_X)}\right)$$

where ρ_X is the correlation coefficient of X within the "fixed effect" group: $\frac{cov(X_{1i}X_{2i})}{\sigma_X^2}$

- – So there is a tradeoff: bias from exclusion of the fixed effect versus bias due to exacerbation of the attenuation in the presence of measurement error with highly correlated X's.

  – Intuition: if the X's are highly correlated, then when using fixed effects, most of the variation left is measurement error.

  – Relavance to matching: One can think of matching as a type of fixed effect. Could be exacerbation of attenuation in the presence of measurement error in treatment. What about measurement error in covariates? Non-classical measurement error? Not yet studied!

- Tradeoff: Matching often allows for better controls, less bias but at the cost of efficiency.

- Many observations thrown away.

- Also since emphasis is on average treatment effect, in the presence of heterogeneity, must use population weights as opposed to weighting be inverse variance: efficiency loss.

- Propensity Score can help when overlap is low (Angrist and Han, 2004)

  * Dont have to throw away observations with low overlap

  * Can gain in efficiency even without gain from less observations thrown out due to greater comparisons per bin

• Previews of Things to Come: Comparison with IV

  - IV estimates ATE if

    1. Homogeneous treatment effect

2. Set of compliers is the entire population

- Matching Estimators estimate ATE if

  1. Homogeneous Treatment Effect

  2. Overlap satisfied at all parts of distribution of covariates (i.e. over full support of covariates)

- Understanding Matching: Future Econometric Research

  - Does matching help with omitted variable bias? No!

  - Can matching help when there is functional form uncertainty and no omitted variable bias (Robustness)? Yes!

  - Can matching help with specification bias? Yes (functional form) and no (variable selection)!

– Unknown: Matching and Measurement Error (In Treatment and in Covariates)

– Unknown: Constructing SEs for Matching Estimators

– Unknown: Balancing Bias (Due to Functional Form) with Efficiency

3 Bootstrapping

- Method for estimating standard errors when techniques don't exist for estimating SEs from econometric theory (for example small sample distributions- i.e. IV), when SE computation is to computationally intensive.

- Conisder

$$Y = XB + \epsilon$$

- Non-Parametric Bootstrapping

  - 1. Estimate true $\hat{\beta}$ from the full sample

    2. Choose N observations at random (with replacement)

    3. Estimate $\hat{\beta}_j$

4. Estimate J of the $\hat{\beta}_j$

5. Either

(a) Test $\hat{\beta}$ relative to the non-parametric distribution of $\hat{\beta}_j$

(b) or compute the variance of the $\hat{\beta}_j$: $\sum_{j \in j} \frac{\left(\bar{\beta} - \hat{\beta}_j\right)^2}{|J|}$ where $\bar{\beta} = \sum_{j \in J} \frac{\hat{\beta}_j}{|J|}$; test using the normality assumption with $\sqrt{V\left(\hat{\beta}_j\right)}$

- Paramtetric Bootstrapping
  - 1. Estimate $\hat{\beta}$ from the full sample

    2. Calculate the residuals: $\epsilon_i = Y_i - X_i\hat{\beta}$

    3. Take the full sample of X_i; for each X_i, resample a residual ϵ_{ij} at random

4. Create a sample of N pairs $\left(X_i, Y_{i_j}\right)$ where $Y_{i_j} = \hat{\beta} X_i + \epsilon_{ij}$

5. Run a regression for each sample and obtain a distribution $\hat{\beta}_m$

6. Either

   (a) Test $\hat{\beta}$ relative to the non-parametric distribution of $\hat{\beta}_m$

   (b) or compute the variance of the $\hat{\beta}_m$: $\sum_{m \in M} \frac{\left(\bar{\beta} - \hat{\beta}_m\right)^2}{|M|}$ where $\bar{\beta} = \sum_{m \in M} \frac{\hat{\beta}_m}{|M|}$ and test using the normality assumption with $\sqrt{V\left(\hat{\beta}_m\right)}$

- Which bootstrap method is preferable?
  - With parametric, you resample ϵ for the different X, which is what you want to do so non-parametric is preferable but:

- If $cov(X, \epsilon) \neq 0$, then your $\hat{\beta}$ which you use to compute Y_{i_j} will be tainted; in this case, it is better to use the parametric bootstrapping

- Block Bootstrap
  - Suppose $cov\left(\epsilon_i, \epsilon_j\right) \neq 0$ for $i \neq j$
  - Then you can block bootstrap (i.e. randomly pick K sequential observations at a time)
  - This way, you randomly sample blocks of data which keeps the error structure in tact

4 Jackknife

- Similar to bootstrap
- Estimate $\hat{\beta}$ from the full sample
- Define $\hat{\beta}_j$ = estimate without the j^{th} observation (could exclude more than one)
- Like boostrapping but
  1. Without replacement
  2. Constructed by excluding variables rather than including them
- Then: either
  1. Test $\hat{\beta}$ relative to the non-parametric distribution of $\hat{\beta}_j$

2. or compute the variance of the $\hat{\beta}_j$: $\sum_{j \in j} \frac{\left(\bar{\beta} - \hat{\beta}_j\right)^2}{|J|}$ where $\bar{\beta} = \sum_{j \in J} \frac{\hat{\beta}_j}{|J|}$; test using the normality assumption with $\sqrt{V\left(\hat{\beta}_j\right)}$

1 Gerber and Green: The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment

- Could ask people at voting polls if they were contacted (in person, by phone or via mail) but selection bias.

- Gerber and Green randomize access to contact (in person, phone or mail). Control is then given by:

$$T_C = \alpha T_R + (1 - \alpha) T_{NR}$$

- Treatment is given by:

$$T_T = \alpha(T_R + t) + (1-\alpha)T_{NR}$$

- The estimate of t (the treatment effect) is then:

$$\begin{aligned} T_t - T_C &= \alpha t \implies \\ t &= \frac{T_t - T_C}{\alpha} \end{aligned}$$

- Randomized contacts independently (mailing, phone, in person contact). Therefore, can look at interactions.
  - Benefits of independent randomization: Can look at interactions between types of contact.
  - Costs of independent randomization: Less power for each type of interaction.

- Implementation
  - Eliminate:
    - Students: why?
    - PO Boxes: why?
  - Mailing: 0, 1, 2, or 3 messages
  - Phone: Survey Company
  - Canvassing: Graduate Students
- What do they estimate:

$$ITT \; : \quad T_C - T_T$$

- IV Estimate:

$$IV \quad : \quad \frac{T_C - T_T}{\alpha}$$

- Computation of α:
  - 25:3% for phone
  - 28:% for visit
  - 12:4% for both
  - for mail?

- Alternative way of writing IV:

$$T_i = \mu + \beta C_i + \epsilon_i$$

$$C_i = \gamma Z_i + \delta_i$$

where T_i is turnout, C_i is contact and Z_i is intention to treat.

- – why is there no constant term in the first stage regression?

TABLE 2. Assignment of Persons to Treatment and Control Conditions

	Number of Direct Mailings Sent				
	None	One	Two	Three	Total
No Telephone Call: Personal canvassing					
No in-person contact attempted	10,800	2,406	2,588	2,375	18,169
In-person contact attempted	2,686	519	625	627	4,457
Telephone Call: Personal canvassing					
No in-person contact attempted	958	1,451	1,486	1,522	5,417
In-person contact attempted	217	385	352	383	1,337
	14,661	4,761	5,051	4,907	
Total					29,380

findings indicate that personal canvassing is highly effective, much more so than the direct mail and telemarketing campaigns that have come to displace it. The implication is that the decline in voter turnout may be due to the changing character of American campaigns. Although the volume of mobilization activity remains considerable, its increasingly impersonal nature draws fewer people to the polls.

EXPERIMENTAL DESIGN

This field experiment was conducted in New Haven, Connecticut, which has a population of approximately 100,000. In September 1998 we obtained a complete list of registered voters, from which we created a data set of all households with one or two registered voters. To eliminate students from the sample, all names with post office box addresses were excluded, as was one voting ward that encompasses a university and student housing. We were left with 29,380 individuals (22,077 households), whose participation in the 1998 election could be determined from public records.

Our study was designed to measure the effect of personal canvassing, telephone calls, and direct mail appeals on voter turnout. Through a series of random assignments, the sample was divided into control and experimental groups. Table 2 shows the sample sizes of each group for the $2 \times 2 \times 4$ design.[3] The treatment and control groups for the three experiments overlap, such that 10,800 people were assigned no intervention; 7,369 were sent at least one mailing but received no other appeal; 2,686 were slated only for personal contact; and 958 were assigned to receive only telephone reminders. The remainder of the sample, 7,567 people, was assigned to two or more treatments. Assignment to the personal canvassing experiment was designed to be uncorrelated with the telephone and mail experiments, so that it could be analyzed separately. Random assignment to each of the telephone/mail treatments was performed in a manner that made calls more frequent among those who received mail. Thus, these two treatments are correlated, and their effects must be estimated using multivariate methods.

Overall, the treatment group for personal canvassing contained 5,794 people, the control group 23,586. For the direct mail experiment, 14,719 people were in the treatment group, and 14,661 were in the control group. The effectiveness of randomization was checked using voter turnout data from the 1996 presidential election. Based on a chi-square test with 15 degrees of freedom for the 16 groups defined in Table 2, we cannot reject the null of independence between treatments and past voting behavior ($p > .10$).

Personal Canvassing Procedure

During each Saturday and Sunday for four weeks before the election, canvassers were sent to contact randomly selected registered voters. They were paid $20 per hour and were primarily graduate students. New Haven has a substantial minority population and a significant proportion of non-English speakers. More than half the canvassers were African American or fluent in Spanish, and when possible they were matched to the racial and ethnic composition of the neighborhoods they walked.

For safety reasons, all canvassing was done in pairs and ceased at sunset. This procedure constrained both the pool of available canvassers and our ability to contact people not at home during the day. In contrast to conventional canvassing efforts, we targeted certain households rather than entire streets, which meant that more time was devoted to locating specific addresses and walking from one to the next. Consequently, canvassers were able to contact only 1,615 (28%) of the 5,794 people in the personal canvassing treatment group.[4] Examination of the data showed a fairly even

[3] Random assignment was done at the household level. The results we present treat individuals as the unit of analysis; as we point out below, however, the results are very similar when we look separately at households containing one or two registered voters. Also, the standard errors we report are very similar to the ones obtained using statistical methods that allow for unmodeled similarities between household members, such as generalized least squares or resampling.

[4] For the subset of persons not contacted, two supplementary experiments were performed. In certain wards, 719 were randomly chosen to receive a mailer, along with a refrigerator magnet that had the election date printed on it. A separate analysis indicated that this

TABLE 4. Effects of Personal Canvassing on Voter Turnout, by Type of Nonpartisan Appeal

Type of Appeal	Turnout Rate	Number of Registered Voters in Treatment Group	Number of Persons Actually Contacted
Unadjusted Turnout Rates among Experimental Subgroups			
Civic duty	47.2%	1,985	534
Neighborhood solidarity	46.3%	1,881	546
Election is close	48.1%	1,928	535
Control	44.8%	23,586	N/A
Implied Effects of Personal Contact on Voter Turnout			
Civic duty	Turnout Differential (2.43%)/Contact Rate (26.90%) = 9.1% Standard Error (4.3)		
Neighborhood solidarity	Turnout Differential (1.48%)/Contact Rate (29.03%) = 5.1% Standard Error (4.1)		
Election is close	Turnout Differential (3.36%)/Contact Rate (27.75%) = 12.1% Standard Error (4.2)		

$$P_E = \alpha(p_r + t) + (1 - \alpha)p_{nr}, \tag{2}$$

where the difference between equations 1 and 2 is due to the effect of the experimental treatment. Combining equations 1 and 2, we derive an expression for t:

$$t = \frac{P_E - P_C}{\alpha}. \tag{3}$$

Although the population probabilities are not observed, sample data can be used to obtain an estimate of t. First, using the law of large numbers,

$$plim\ V_E = P_E, \qquad plim\ V_C = P_C, \tag{4}$$

where V_E is the percentage of the treatment group that votes, and V_C is the percentage of the control group that votes. Similarly,

$$\frac{N_r}{N_E} = \alpha, \tag{5}$$

where N_r is the number of subjects in the treatment group who were reached for the experimental treatment, and N_E is the number of subjects in the treatment group overall. Using equations 4 and 5, we obtain a consistent estimator of t:

$$plim \frac{V_E - V_C}{\dfrac{N_r}{N_E}} = t. \tag{6}$$

Equation 6 says that, to find the treatment effect, subtract the turnout rate of the control group from the turnout rate of the experimental group and divide this difference by the observed "contact rate," which is 28%. Using this formula, we find that personal contact raises the probability of turnout by 8.7 percentage points, with a standard error of 2.6. The null hypothesis that canvassing does nothing to increase turnout can be decisively rejected, using a one-tailed test ($p < .01$).[8]

Table 4 suggests that the effects of personal contact do not vary significantly across messages. The close election message boosts turnout rates by 12.1%, which is slightly better than the 9.1% associated with the civic duty appeal and substantially better than the 5.1% for neighborhood solidarity. These findings are suggestive, but the standard errors associated with the estimates are far too large to reject the null hypothesis that the messages have equal effects. Looking ahead to the experiments using direct mail and telephone calls, we find a similar pattern of insignificant differences across messages. Since we cannot rule out the view that any plausible mobilization appeal works equally well, the analysis that follows focuses exclusively on the relative effectiveness of delivering the appeal in person, by telephone, or through the mail.

Regression Results

Regression analysis permits us to conduct a more comprehensive analysis, taking into account all the treatments in our experiment. Regression analysis has the further virtue of introducing covariates, such as past voting history, that reduce the unexplained variance in voting rates and allow for more efficient estimation of the experimental effects. For reasons cited above, however, any regression analysis must attend to the possibility that subjects with a higher propensity to vote are easier to reach in person.

Consider the following simple model of how the experimental treatment affects turnout. Suppose again, for purposes of illustration, that the population can be divided into those who are easy to contact and those who

[8] These results remain unchanged when we disaggregate the data according to whether the household contains one or two registered voters. For single-voter households, the effect of personal contact is estimated to be 10.0% (SE = 3.7), compared to 8.2% (SE = 3.6) for two-voter households. These estimates are too similar to be differentiated statistically ($p > .10$).

TABLE 5. Linear and Nonlinear Regression of Voter Turnout on Mode of Contact, with and without Covariates

	Two-Stage Least Squares		Two-Stage Probit
Independent Variables	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Personal contact	.087 (.026)	.098 (.022)	.323 (.074)
Direct mailings (0 to 3)	.0058 (.0027)	.0063 (.0023)	.0214 (.0067)
Telephone contact	−.047 (.023)	−.035 (.020)	−.130 (.056)
Registered as Democrat or Republican		.064 (.006)	.217 (.015)
Voted in 1996 general election		.229 (.007)	.589 (.018)
Abstained in 1996 general election		−.231 (.008)	−.824 (.024)
Age		.0188 (.0008)	.0649 (.0022)
Age squared		−.000133 (.000007)	−.000467 (.000020)
Number of registered voters in household (1 or 2)		.056 (.005)	.188 (.014)
Constant	.445		
F	5.86	296.66	
Degrees of freedom	29,376	29,342	29,342

Note: The base category for past voting behavior is the set of people who were not registered in 1996. Not reported in this table are the coefficients associated with each of the 29 wards. The first-stage equations include dummy variables representing the intent-to-treat groups associated with canvassing, phone calls, and direct mail. The first-stage equation also includes covariates for columns 2 and 3. Standard errors for the two-stage probit estimates were obtained using jackknifing.

are not. The probability that a given person in the experiment votes may be expressed as

$$Y = a + b_1X_1 + b_2X_2 + e,$$

where $Y = 1$ if the subject votes, $X_1 = 1$ if the subject is difficult to contact, and $X_2 = 1$ if the subject is actually contacted; 0 otherwise. Given that X_1 is not observed, we might ignore this variable and regress Y on an intercept and X_2. This will yield a consistent regression coefficient estimate only if X_1 and X_2 are uncorrelated, or if b_1 equals 0. These special conditions cannot be expected to hold. Unless everyone in the treatment group is contacted, there will be some correlation between how easy it is to reach a subject and the likelihood they are actually reached. It is also quite reasonable to assume that those who are very hard to reach may also be less likely to vote (i.e., b_1 does not equal 0). Although these points seem straightforward, they have eluded previous research in this area.[9]

The standard solution to the problem of correlation between a right-hand-side variable and the regression error is to find a suitable instrumental variable. In this case, an ideal instrument is at hand. Recall that a valid instrument satisfies two criteria: The variable must be uncorrelated with the regression error, and it must be correlated with the endogenous variable. The probability that subjects are contacted is a function of whether they are randomly selected for the treatment group. This implies that a dummy variable which equals 1 for subjects in the treatment group will be correlated with the endogenous variable. Because the treatment group is generated through random assignment, there is no reason to suppose that those who are easy to contact will be overrepresented. Thus, the expected correlation between the instrumental variable and the regression error is zero.

Table 5 presents two-stage least-squares regression estimates of the effect of each experimental treatment. As indicated earlier, the instrumental variables used in the regressions indicate whether the person was in a given treatment group. For example, the variable *Personal Contact* equals 1 if the subject was contacted, and the instrumental variable equals 1 if the person was in the group that we intended to treat. Note that the instrumental variable will be correlated with the included variable (being in the intent-to-treat group predicts the likelihood that one is contacted), but the instrumental variable is not correlated with the regression error (treatment group status is due to random assignment). A similar procedure applies to the telephone experiment, with intent-to-treat serving as an instrument for actual contact. For the mail experiment, the instrumental variable and the independent variable are the same, since the assumed contact rate is 100%.[10]

Official voting and registration records contain useful information about the sample. For example, we know whether a person voted, abstained, or was absent from the voter rolls in the 1996 general election. We also know an individual's age, party registration, voting ward, and whether s/he is the sole registered adult in the household or is one of two. Each of these covariates contributes significantly to the predictive accuracy

[9] Consider some of the seminal work in this area. Kramer (1970) interprets the higher turnout rate among those reached by a party or candidate as the marginal effect of contact. In the classic study by Eldersveld (1956), those unavailable for personal contact were moved into the control group. This practice results in overestimation of the treatment effect.

[10] Our calculations assume that all of the households we intended to treat by mail received the treatment, an assumption implicitly made in all previous mail experiments. In our case, the voter lists were very current and fewer than 1% of the mailings were returned. To adjust the estimated effects for any failure to receive the mail, divide the coefficients in Table 5 by the supposed contact rate.

2 Imai: Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments

- Gerber/Green experiment not properly done; covariates are not balanced across assignment to treatment and to control

- Use logistic regression of assignment to treatment on covariates

$$P(T_i) = F(X_i\beta) + \epsilon_i$$

- Construct fitted probabilities for each treatment and control and use nearest neighbor matching:

- For each treatment with propensity $\hat{p}\left(T_i^T\right)$, choose the nearest control: $\hat{p}\left(T_{j*}^C\right)$ s.t. $\hat{p}\left(T_{j*}^C\right) = \min_{\hat{p}\left(T_j^C\right)} \left|\hat{p}\left(T_i^T\right) - \hat{p}\left(T_j^C\right)\right|$
- If more than one control minimizes the nearest neighbor criterion, then randomize

- Then check covariate balance

- Add higher order terms in logistic regression if covariate balance has not been acheived (including interactions)
  - Dont report final logistic regression equation
  - Is adding higher order terms and interactions until you get covariate balance cheating?

- Variables:

- Age
- Voted in 96'
- New Registered Voter
- Registered Democrat
- Registered Republican
- Two-voter household
- Ward of residence

- Then recompute ITT effects (not weighted to population):

$$\sum_{t=1}^{T} \left(Y_t - Y_{M(t)} \right)$$

- Standard errors are bootstrapped with 500 replications candidates may want to know about how many visits or postcards are necessary to increase voter turnout by one percentage point. In this case, it is not necessary to know how many voters actually talked to canvassers or read postcards. On the other hand, political scientists, who want to assess the relative effectiveness of various canvassing methods need this extra information. Even when personal canvassing seems less effective, for example, it may only appear ineffective because voters are more difficult to reach by visits than by postcards. Hence, the different compliance rates for the two methods become critical.

THE NEW HAVEN VOTER MOBILIZATION STUDY

In this section, I replicate and extend Gerber and Green's analysis of the voter mobilization study. Gerber and Green (2000) designed and conducted an experiment where registered voters in randomly selected households of New Haven were encouraged to vote in the 1998 general election by means of personal visits, phone calls, and postcards. They then examined voting records and analyzed which strategies had increased voter turnout. In addition to the voting record of the 1998 election, the data include covariates that describe the following characteristics of each registered voter: number of registered voters in the household (one or two), age, party affiliation (registered Democrats, registered Republicans, or others), voting record in the last general election (voted, did not vote, or was not registered for 1996 election), and ward of residence in New Haven (29 wards).

Inefficient Experimental Design

Table 1 shows the unusually complicated experimental design of the original study with the substantial overlap of different treatment assignments. Over 40% of voters in the sample were assigned more than one treatment. For example, 122 voters were assigned to receive three postcards, a phone call, and a personal visit with the civic duty message. Further variation in the nature of the treatment was possible because Gerber and Green used three different appeal messages; civic duty, neighborhood solidarity, and close election. The authors note that the neighborhood solidarity message was not used for phone calls (Gerber and Green 2000, 656). Altogether, this design produced a total of 45 different treatment combinations and their corresponding potential outcomes.

Such complex experimental design leads to the inefficient estimation of treatment effects unless one makes arbitrary assumptions. This is unfortunate since the advantage of experimental methods is to avoid additional assumptions that are often necessary in observational studies. For example, Gerber and Green (2000) assume that the effect of telephone canvassing remains the same regardless of whether voters have received other treatments. However, phone calls may not increase the probability of voting as much for those voters who already have received a personal visit. Furthermore, the timing of contact differs from one canvassing method to another and this variation was not randomized; e.g., phone calls were made during the three days prior to the election, whereas personal visits were made over a period of four weeks. Such systematic differences in the administration of multiple treatments will yield incorrect inferences unless properly controlled in the analysis.

TABLE 1. The Original Experimental Design Reported in Gerber and Green (2000)

	Mail			
	None	Once	Twice	3 times
Phone				
Visit				
Civic	33	103	126	122
Neighbor/civic[a]	74	144	113	127
Close	110	138	113	134
No visit				
Civic	<u>581</u>	443	432	479
Neighbor/civic[a]	0	491	520	542
Close	<u>377</u>	517	534	501
No phone				
Visit				
Civic	<u>1,011</u>	150	213	227
Neighbor	<u>853</u>	175	201	194
Close	<u>822</u>	194	211	206
No visit				
Civic		<u>870</u>	<u>922</u>	<u>825</u>
Neighbor	$\boxed{10,800}$	<u>764</u>	<u>849</u>	<u>767</u>
Close		<u>722</u>	<u>817</u>	<u>783</u>

Note: The figures represent the number of registered voters in New Haven for each treatment assignment combination. For example, 122 voters were assigned to receive three postcards, a phone call, and a personal visit with the civic duty message. Treatment assignment groups of interest are underlined. A box highlights the large control group.
[a] For phone calls, the civic duty appeal was used instead of the neighborhood solidarity message (Gerber and Green 2000, 656).

Incorrectly Identified Treatment Assignment and Control Groups

Gerber and Green (2000) also incorrectly identified the treatment assignment and control groups used in their field experiment and, as such, failed to estimate their causal quantities of interest. For example, when estimating the marginal effect of phone calls, Gerber and Green used the treatment assignment group that includes those who were also assigned other treatments such as personal visits and postcards (the upper two rows in Table 1). Their control group included those voters who were assigned other treatments (all categories in the bottom two rows in Table 1). In order to correctly estimate the treatment and ITT effects, the appropriate control group should consist solely of the 10,800 voters who were assigned *no* treatment and hence received no intervention. Likewise, the members of the treatment assignment group for phone calls should not include those who were assigned any other treatment.

TABLE 2. Treatment Assignment and Control Groups Based on the Revised Data

	Mail			
	None	Once	Twice	3 times
Phone				
Visit				
Civic	0	88	107	98
Civic/blood[a]	104	17	21	17
Civic/blood-civic[b]	0	12	9	18
Neighbor	0	109	92	101
Neighbor/civic[c]	74	22	15	15
Neighbor/civic-neighbor[d]	0	13	6	11
Close	110	138	113	134
No visit				
Civic	428	385	352	411
Civic/blood[a]	371	84	98	95
Civic/blood-civic[b]	0	29	46	33
Neighbor	0	374	367	390
Neighbor/civic[c]	0	73	102	97
Neighbor/civic-neighbor[d]	0	44	51	55
Close	377	517	534	501
No phone				
Visit				
Civic	940	136	202	216
Neighbor	853	175	201	194
Close	822	194	211	206
No visit				
Civic		815	858	765
Neighbor	10,582	764	849	767
Close		772	817	783

Note: The figures represent the number of registered voters in New Haven for each treatment assignment combination. For example, 104 voters were assigned a phone call with the blood donation message and a personal visit with the civic duty appeal. Treatment assignment groups of interest are underlined. A box highlights the control group.
[a] For phone calls, the blood donation appeal was used instead of the civic duty message.
[b] For phone calls, either the blood donation or the civic duty appeal was used.
[c] For phone calls, the civic duty appeal was used instead of the neighborhood solidarity message.
[d] For phone calls, either the civic duty or the neighborhood solidarity appeal was used.

This implies that the ITT and treatment effects reported in Gerber and Green (2000) are confounded by the effects of other treatments.[5] In experiments, an appropriate control group is critical to ensure internal validity (e.g., Campbell and Stanley 1963). In principle, it is advisable to minimize the number of treatments in field experiments. Although factorial designs may be feasible in laboratory experiments, additional complications such as noncompliance make it difficult to estimate the effects of multiple overlapping treatments in field experiments. In this article, I focus on the marginal effects of each treatment rather than their interaction effect, as the latter would involve additional assumptions and few data are available to estimate such quantities.

[5] This may lead to the underestimation of the treatment effect since the control group used by Gerber and Green includes those who received other treatments. Many voters in the treatment assignment group were also assigned one or more of the other treatments. The treatment effects are likely to be small for those who have already received other treatments.

Experimental Design Based on the Revised Data

As noted above, the analysis in the initial draft of this article detected the implementation errors and led to the subsequent revisions of the original data. Table 2 shows the treatment assignment and control groups based on the most recent data and Gerber and Green's latest version of their experimental design. The total number of treatment combinations is now seventy, making the experimental design even more complex. For the analysis of the revised data, I correct the treatment group for telephone canvassing to include only those voters who were assigned no other treatment. I also exclude those who were possibly assigned the blood donation messages. This yields the total of 428 voters with the civic duty appeal and 377 individuals with the close race message. The new control group consists of 10,582 voters who were assigned no treatment.

The analysis of the revised data reveals discrepancies between Gerber and Green's description of the implementation errors and the altered coding scheme.

TABLE 3. Estimated Average Intention-To-Treat (ITT) Effects on Voter Turnout Assuming Complete Randomization (Percentage Points)

	Original Data		Revised Data
Treatment	Gerber & Green (Incorrect Groups)	Corrected ITT (Correct Groups)	(Correct Groups)
Phone[a]	−1.5	−2.9	−0.9
	(0.7)	(1.7)	(1.8)
Visit	2.4	3.9	3.6
	(0.7)	(1.1)	(1.1)
Mail			
Once	0.6	0.4	0.5
	(0.3)	(1.1)	(1.1)
Twice	1.2	0.8	0.8
	(0.5)	(1.1)	(1.1)
3 times	1.7	2.6	2.7
	(0.8)	(1.1)	(1.1)

Note: The left column of estimates displays the results based on the incorrectly identified groups as published in Gerber and Green (2000). The ITT estimates in the middle column use the proper treatment assignment and control groups, thereby correcting the original analysis of Gerber and Green (2000). Finally, the estimates in the right column are based on the revised data using the correct treatment assignment and control groups. Standard errors are in parentheses.
[a] The ITT effect of phone calls was not reported by Gerber and Green (2000) and is calculated based on their method.

For example, on their Web site they describe one of their errors as follows: "Subjects who would have received Civic Duty *mail* or *personal* appeals received *phone* appeals requesting a Blood Donation" (see footnote 4). Although this error should not affect the control group of those who were assigned no treatment in the first place, the revised control group has about 300 voters fewer than the original group. Such remaining inconsistency calls for further clarifications about the coding changes beyond what is currently documented.

ANALYSIS ASSUMING COMPLETE RANDOMIZATION WITH CORRECTED TREATMENT ASSIGNMENT AND CONTROL GROUPS

With the corrected treatment assignment and control groups, I reestimate the average ITT and treatment effects by applying the statistical method used in Gerber and Green (2000), which assumes complete randomization of treatment assignments.

Estimation of the ITT Effect

Under the assumption of complete randomization, the treatment assignment is independent of all observed and unobserved individual characteristics. Therefore, the difference in the sample means of the treatment assignment and control groups is an unbiased estimate of the average ITT effect. Namely,

$$\widehat{ITT} = \frac{\sum_{i=1}^{N} Y_i Z_i}{N_1} - \frac{\sum_{i=1}^{N} Y_i (1 - Z_i)}{N_0}, \qquad \textbf{(3)}$$

where $N_1 = \sum_{i=1}^{N} Z_i$ is the size of the treatment assignment group $N_0 = \sum_{i=1}^{N}(1 - Z_i)$ is the size of the control group, and $N = N_0 + N_1$.[6]

Table 3 shows the results of the ITT analysis using the correct treatment and control groups. First, the corrected ITT analysis in the middle column confirms the conclusion of Gerber and Green (2000) that personal canvassing is the most effective method for increasing voter turnout. Second, get-out-the-vote calls have a significant negative effect on turnout. Using the appropriate treatment assignment and control groups does not change the odd finding of the original article that telephone canvassing reduces voter turnout.

As one would expect, altering the data also changes the estimates. The analysis of the revised data with correct groups (in the right column) suggests that the overall ITT effect of phone calls is only slightly negative, with a larger standard error. In the next section, however, I show that the data correction alone does not solve the entire problem. In principle, the implementation errors of field experiments cannot be fixed by the experimenter after the fact without statistical adjustments.

Mail canvassing also mobilizes voters. (Gerber and Green 2000, 661) argued that "even if the effective marginal costs of canvassing were doubled, face-to-face mobilization would still be cost effective." This conclusion, however, is based on their assumption that all voters who were sent postcards actually received and read them (659, fn 10). Such an assumption is not warranted because many cards may not have reached a voter due to changes of address or may have been

[6] In the case of phone calls, for example, $N_1 = 958$ and $N_0 = 10,800$.

TABLE 6. Probability of Successful Randomization with Respect to Observed Covariates in Gerber and Green's Field Experiment

	Original Data		Revised Data	
Treatment	Probability	N	Probability	N
Phone	0.035	958	0.0085	805
Visit	0.000012	2,686	0.0000098	2,615
Mail	0.0000000035	7,369	0.00000000054	7,190

Note: Probability represents the p-value of the residual deviance test from a logistic regression model predicting the assignment of each treatment given all observed covariates and their first-order interactions. N represents the size of the treatment assignment group. The last row in the second column, for example, tells us that under the assumption of successful randomization, the pattern of incomplete randomization for mailings observed in Gerber and Green's original data would occur only with a probability of about one in 300 million. These probabilities cannot be compared across different treatments because of different sample sizes.

randomization observed in Gerber and Green's data can occur only with a probability of one in 300 million. This probability is smaller for the revised data, reaching to one in 2 billion. (Note that a small sample size makes it harder to detect failure of randomization, so that the larger p-value for phone calls than for visits and mailings does not necessarily imply that randomization was more successful.) In sum, the test with respect to observed covariates also provides strong evidence that treatment assignment was not randomized in Gerber and Green's field experiment.

In field experiments, randomization of treatment assignment is not as easy to accomplish as one might expect. In practice, it is often difficult to randomize every aspect of each treatment. In Gerber and Green's experiment, personal canvassing was conducted over a period of four weeks before the election, whereas telephone canvassing took place over three days including the election day. Postcards were sent out during the two weeks before the election. Although a visit right before the election would have a greater effect than a visit one month before the election day, the timing of contact was not randomized. Likewise, the effect of different canvassers, if not randomized, can confound the effect of different canvassing methods. These examples illustrate the difficulty of randomization and potential confounding effects that threaten the validity of field experiments.

Finally, I investigate the sources of the negative finding about phone calls. Both Gerber and Green's analysis and the corrected IV analysis indicate that telephone canvassing has a large and negative effect on voter turnout among single-voter households. I find that for this subgroup the assignment of phone calls was not randomized with respect to the past voting record. In particular, only 42% of the treatment assignment group voted in the last election, whereas 47% of the control group voted (p-value, 0.05). The randomization for this group appears to be incomplete even with the incorrectly identified treatment assignment and control groups used by Gerber and Green.[17] Since those who voted in the last election are 40 percentage points more likely to vote in the current election on average, this difference contributes to the large negative effects of phone calls for single-voter households.

When One Should Not Use IV Estimation

The large bias of IV estimation that results from violation of the exclusion restriction is well documented (e.g., Angrist, Imbens, and Rubin 1996, 450). In particular, the bias is worsened when unbalanced variables are good predictors of the outcome variable and when a large number of noncompliers exist. Equation (4) illustrates these two conditions; the bias of the IV estimate is large (a) when the bias of the ITT estimate due to incomplete randomization is large and (b) when the compliance rate is low. (Recall that the IV estimate is equal to the ITT estimate divided by the estimated compliance rate.)

Gerber and Green's study fits both conditions for large bias. First, the unbalanced covariates (i.e., the voting record in the previous election) predict turnout well, which suggests that the bias in the estimated ITT effect is large. Furthermore, the compliance rate of this field experiment is low (about 25% for phone calls). This low compliance rate implies that if the ITT effect is biased by five percentage points, for example, then the bias of the IV estimate can be as large as 20 percentage points. Thus, the combination of a large bias in the ITT estimate and low compliance rate led to the puzzling finding that get-out-the-vote calls significantly decrease turnout.[18]

If one successfully randomizes the treatment assignment, the method of instrumental variables can give estimated treatment effects that are consistent in large samples. However, as the analysis of this section suggests, making this assumption in practice requires careful experimental design and successful implementation. In this case, the failure of randomization for telephone canvassing led to inaccurate causal inferences

[17] Compared with the control group, the treatment assignment group includes significantly more individuals who abstained in the last election. The mean difference is statistically significant at the 0.05 level.

[18] It is also important to note the finite sample bias and inefficiency of IV estimation (e.g., Bound, Jaeger, and Baker 1995). The small size of each treatment group in the New Haven mobilization study suggests the importance of finite sample consideration.

TABLE 7. Differences in Observed Characteristics between Compliers and Control Group Prior to Matching Adjustment

	Phone Call			Personal Visit		
Variable	Mean Diff.	*t* Stat.	Var. Ratio	Mean Diff.	*t* Stat.	Var. Ratio
Age	9.01	7.00	1.12	3.22	4.66	0.96
Voted in '96 election	18.8%	6.41	0.81	3.9%	2.10	0.99
Newly registered voter	−8.9%	−4.32	0.62	−0.5%	−0.33	0.98
Registered Democrat	5.5%	1.95	0.89	3.0%	1.76	0.94
Registered Republican	0.6%	0.40	1.11	−1.2%	−1.55	0.80
Two-voter household	2.6%	0.79	1.00	−0.3%	−0.17	1.00

Note: The table shows the differences in covariate distributions due to noncompliance. The mean of each covariate for the control group is subtracted from that for the treatment group. The *t* statistics for these mean differences are also reported. The variance ratios are calculated by dividing the variance of the treatment group by that for the control group. Matching would be unnecessary if mean differences were near zero and the variance ratios were near one.

about the effects of get-out-the-vote calls in Gerber and Green (2000).

ANALYSIS WITHOUT ASSUMING COMPLETE RANDOMIZATION

The previous section showed that IV estimation was inappropriate for telephone canvassing given the incomplete randomization of treatment assignment. This calls for more general statistical methods to estimate the effects of nonrandom treatments. I apply propensity score matching to reduce the bias caused by nonrandom treatment.[19] Matching is particularly useful for field experiments when randomization of treatment assignment is incomplete and important covariates are available. The basic idea of matching follows the logic of causal inference described earlier. The goal is to construct a control group as similar to the treatment group as possible. The method of matching finds two groups of subjects who have exactly the same observed characteristics except that one receives the treatment and the other does not. Since matching is a nonparametric method, it does not require the assumptions of usual regression analysis, (e.g., linearity and additivity), and hence it effectively reduces bias due to incomplete randomization.

The intuition behind matching resembles the traditional comparative case study method, which dates back to John Stuart Mill (1930/1843). Both approaches call for comparing cases that are very similar to each other except for the primary causal variable. This facilitates the evaluation of main causal effects in isolation by reducing the possibility of confounding effects from other variables. Although the comparative method has largely been used for qualitative studies, with the method of matching, quantitative and historical case studies can rest on a common ground of causal inference.

Selection Bias Due to Noncompliance

In field experiments, even when treatment assignment is completely randomized, the actual treatment group of compliers ($T_i = Z_i = 1$), as opposed to the treatment assignment group ($Z_i = 1$), is often different from the control group ($T_i = Z_i = 0$) in its characteristics. Table 7 illustrates the imbalance of observed covariates between compliers and the control group. The wide gap between the two groups indicates a significant selection bias that requires statistical adjustment. Compliers are older, are more Democratic, and have a better past voting record than the control group. A similar pattern is observed in the revised data. Estimates of treatment effects will be biased, unless one properly adjusts for these systematic differences between the two groups. Next, I explain how propensity score matching effectively reduces this selection bias.

Matching

The key assumption of matching is that compliers in the control group can be identified using their observed characteristics. In other words, the assumption implies that it is possible to estimate the counterfactual outcome under no treatment for a treated unit by using individuals from the control group who share the same observed characteristics. Formally, the counterfactual outcome under no treatment, $Y(T = 0)$, is assumed to be mean independent of the actual treatment status, T, conditioning on the set of observed control variables, X (e.g., Heckman et al. 1998),

$$E\{Y(T = 0) \mid T = 1, X\} = E\{Y(T = 0) \mid T = 0, X\}. \quad \textbf{(5)}$$

Equation (5) implies that matching effectively reduces bias when important covariates are observed. Omitted variable bias is possible if the observed covariates, X in Equation (5), do not contain variables that affect both T and $Y(T = 0)$. The bias can be reduced,

[19] The estimand for the method of matching (i.e., the average treatment effect for the treated) can differ from that for IV estimation (i.e., the average treatment effect for compliers). In the New Haven mobilization study, however, the two estimands are equivalent because the treated did not include "always-takers," who take the treatment even when they are not assigned the treatment (i.e., it is assumed that $T_i = 0$ if $Z_i = 0$). See Angrist, Imbens, and Rubin 1996 for a complete discussion of this issue.

TABLE 8. Similarity of Observed Covariates between Treatment Compliers and Matched Control Groups

	Phone Call			Personal Visit		
Variable	Mean Diff.	*t* Stat.	Var. Ratio	Mean Diff.	*t* Stat.	Var. Ratio
Age	0.23	0.17	0.97	0.16	0.21	1.00
Voted in '96 election	−0.8%	−0.26	1.02	−0.1%	−0.06	1.00
New registered voter	−1.0%	−0.44	0.93	−0.3%	−0.16	0.99
Registered Democrat	1.4%	0.45	0.97	−1.1%	−0.61	1.03
Registered Republican	−0.2%	−0.14	0.97	0.3%	0.36	1.07
Two-voter household	1.9%	0.54	1.00	0.2%	0.08	1.00
Ward of residence	25.5% matched			35.4% matched		
Exact match	19.3% matched			25.7% matched		

Note: The table shows that matching effectively balances the observed covariates. The mean of each covariate for the control group is subtracted from that for the treatment group. The *t* statistics for these mean differences are also reported. The variance ratios are calculated by dividing the variance of the treatment group by that for the control group. Compared with Table 7, the mean differences are closer to zero and the variance ratios are closer to one, indicating that the covariate balance of the two groups is significantly improved by matching.

If there is more than one voter with the same propensity score, I randomly select one of them. I repeat this procedure to obtain several matched control units for each treated unit; five matches for phone calls, yielding 1,210 selected control units, and three matches for personal visits and mailings (three postcards), yielding 2,268 and 7,125 matched control units, respectively. Increasing the number of matched control units generally improves the efficiency of resulting estimates because more observations are included in the analysis, but it will typically produce a greater imbalance of covariates between treated and matched control units, which in turn may lead to biased estimates. As shown below, different matching schemes can also be used for sensitivity analysis to detect this potential bias.

To estimate the propensity score, I use logistic regression starting with the specification where I include all available covariates as linear predictors. When this model does not balance all covariates, I search for an alternative specification by including additional terms to improve the balance.[21] I use mean differences and variance ratios to investigate the resulting balance of covariates and determine model specification. Since all covariates except age of voters are indicator variables, these two statistics are generally sufficient to measure the similarity of the covariate distributions between treated and matched control units. The availability of such diagnostic tests for model specification is an important advantage of propensity score methods.

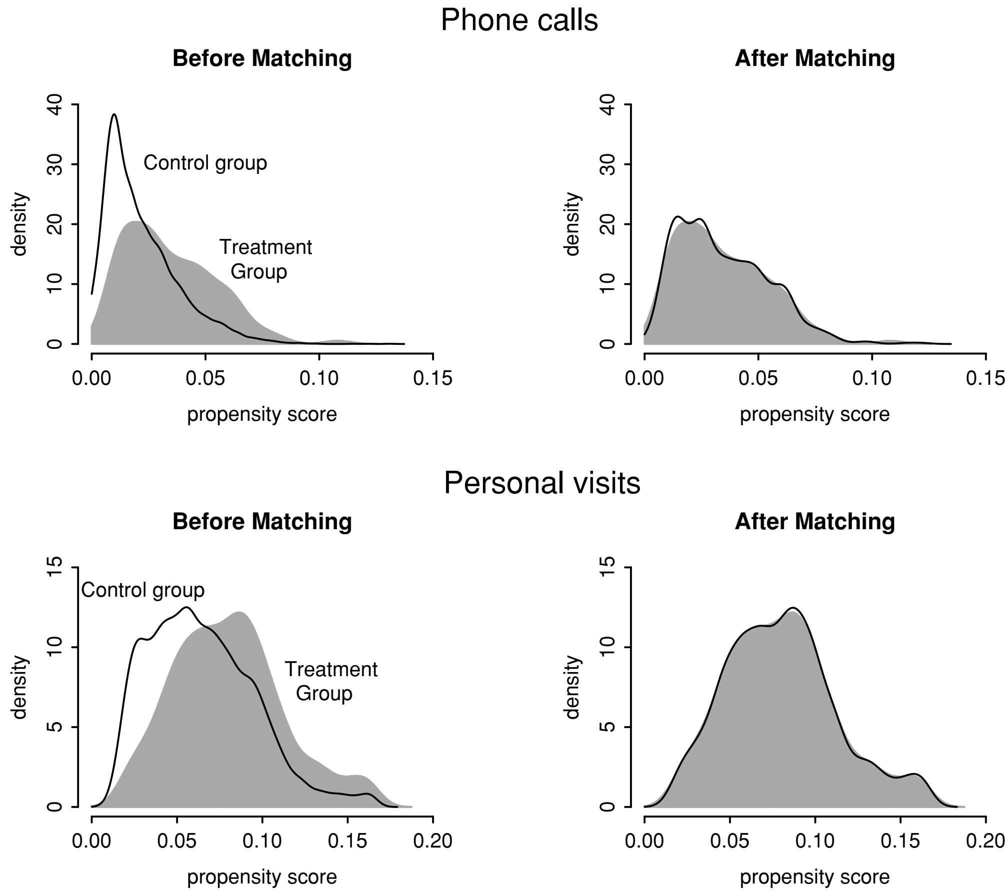
Table 8 shows that matching on the estimated propensity score successfully balances all observed covariates. The mean differences of all covariates between the treated units and the control-group individuals are not statistically significant and their variances are similar. In particular, propensity score matching significantly improves the balance of covariates compared with Table 7. I also find many exact matches. For phone calls, about one fifth of the matched control units share exactly the same values of all covariates with one of the treated units. That is, they live in a household with the same number of registered voters, are exactly the same age, have the same party affiliation, reside in the same ward of New Haven, and have the same voting record in the previous election. Similarly, in the case of personal visits, I find about one fourth of the matched control units to be exact matches.

Figure 2 further compares the similarity of the two groups by examining the distributions of the estimated propensity score. Since the propensity score is a scalar summary of all observed covariates, successful matching should produce a matched control group whose propensity score distribution is similar to that of the treatment group. While the distributions of the treatment group (indicated by the gray density) and control-group individuals (indicated by the solid line) are substantially different before matching, they are almost identical after matching.

Finally, the same test as shown in Table 6 can be applied to the matched sample. I use the same logistic regression to predict the receipt of each treatment in the sample that combines those who received the treatment with a group of compliers selected by matching. If matching is successful, the model should not predict the receipt of any particular treatment well. The results show that after matching, the model no longer predicts the receipt of treatments. Indeed, using the original data, the p-values for phone calls, personal visits, and postcard mailings are 0.63, 0.67, and 0.65, respectively. For the revised data, the results are 0.84, 0.88, and 0.99. The large p-values contrast with the

[21] The model specifications for the original data are as follows. For phone calls, the household type variable is interacted with past voting record. For personal visits, the household type is interacted with the other variables except the new voter variable. Both models include the square term of age. For mailings, the household type is interacted with age, past voting record, and ward of residence variables. The model specifications for the revised data are as follows. For phone calls, the square term of age and the two interaction terms of the household type, one with the past voting record and the other with the new voter variable, are added. For personal visits, the interaction terms of the household type with the other variables except the past voting record are added. For mailings, the household type was matched first, and all first-order interaction terms are included.

FIGURE 2. Distributions of Propensity Scores for Treatment and Control Group Before and After Matching Adjustment



Note: The graphs are smooth versions of histograms produced with Gaussian kernels. Gray areas and solid lines represent the distributions of propensity scores for treatment and control groups, respectively. Before matching adjustment, the two distributions are quite different (left). After matching, however, they are almost identical (right).

results in Table 6, confirming that the matched sample balances the covariates between the treatment and the control groups.

The effectiveness of matching illustrates an important advantage of randomized field experiments. In many observational studies, it is often difficult to conduct matching adjustment because the treatment group is too different from the control group. For such cases, even the propensity score may prove inadequate. In field experiments, such problems are less likely because the control group tends to be a representative sample of the relevant population. Despite the randomization problems for phone calls, Gerber and Green's study produced treatment assignment and large control groups for which propensity score matching can effectively balance all covariates.

GET-OUT-THE-VOTE CALLS INCREASE TURNOUT

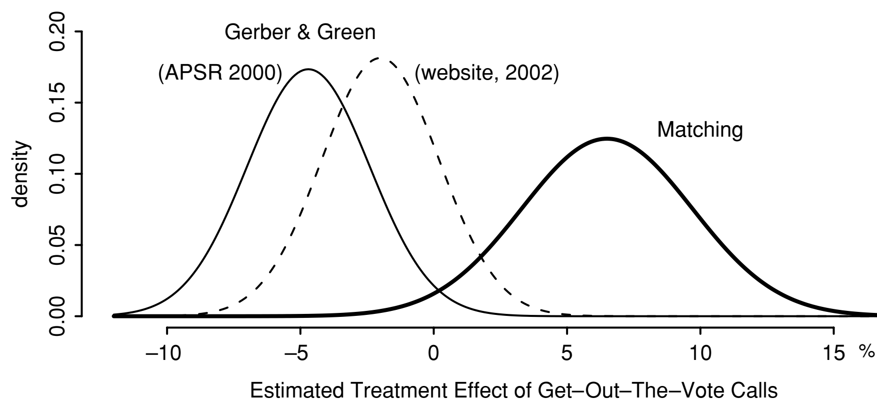
After matching with the estimated propensity score, I calculate the average treatment effects of phone calls and personal canvassing as well as the average ITT effects of mailings (three postcards). Table 9 presents the matching estimates for revised data. The results based

TABLE 9. Matching Estimates of Average ITT and Treatment Effects on Voter Turnout (Percentage Points)

	Phone	Visit	Mail
Overall effect	6.5	9.2	1.5
	(3.2)	(2.1)	(1.0)
Single-voter households	6.9	9.6	0.7
	(4.8)	(3.1)	(1.7)
Two-voter households	6.1	8.9	2.2
	(4.7)	(2.9)	(1.2)

Note: The average treatment effects are estimated for personal visits and phone calls, while the average ITT effects are estimated for mail canvassing (three postcards). The results are based on 500 bootstrap replications. Standard errors are in parentheses.

FIGURE 3. Comparison of Matching Estimates and Gerber and Green's Results for Average Treatment Effect of Get-Out-the-Vote Calls



Note: The estimated average treatment effect of phone calls. The Normal distribution is used to approximate the distribution of estimates. While the matching estimates indicate that phone calls have a positive impact on turnout, Gerber and Green's results (*APSR* 2000, solid line; Web site 2002, dashed line) imply otherwise.

on the original data are similar.[22] The results show that get-out-the-vote calls *increase* turnout by a little more than six percentage points on average (with a standard error of 3 percentage points), reversing the negative finding reported in Gerber and Green (2000). While it may not appear as effective as personal visits, telephone canvassing offers a significant alternative mobilization strategy. The matching estimate for personal visits is significantly smaller than the corrected IV estimate. The estimated ITT effect of sending three postcards is about two percentage points. Mailing appears to be especially effective for two person households, suggesting that along with phone calls, mail canvassing may represent another cost-effective mobilization strategy.

Although the overall effect of postcards may appear to be smaller than that of phone calls and visits, such a simple comparison is misleading. While the ITT effect of postcards is estimated for the entire treatment assignment group, the treatment effects of the subgroup of compliers are estimated for the other canvassing methods. In particular, it is possible that postcards may be as effective for compliers as phone calls and visits are for this subgroup. Unless we have the information about who actually read postcards, it is difficult to identify the treatment effect of postcards for compliers.

Figure 3 compares the matching estimates with the original results reported in Gerber and Green (2000) as well as the estimates posted on their Web site (see footnote 4). (When analyzing the revised data, Gerber and Green incorrectly identify their treatment and control groups. Thus, their estimates differ from the corrected IV estimates reported in Table 5, which are based on the actual treatment assignment and control groups.) The conclusions one would draw from two statistical methods are clearly different. Matching shows that get-out-the-vote calls increase turnout, whereas Gerber and Green's IV analysis indicates that such calls may discourage voters from casting their ballots. Although Gerber and Green's Web site results are somewhat closer to my matching estimates, the difference shows that the data correction alone is not sufficient to fix all the problems that have occurred when implementing their field experiment.

The positive finding about telephone canvassing agrees with the results of another experimental study recently conducted in a different setting by the same authors as well as the earlier experimental results (e.g., Adams and Smith 1980, Eldersveld 1956, and Miller, Bositis, and Baer 1981). In their recent study, Green and Gerber (2001, 2) conclude that "phone canvassing increased turnout by an average of 5 percentage-points. This finding, based on six experiments involving nearly 10,000 people, is statistically significant."[23] Given that making a phone call costs much less than visiting a home, get-out-the-vote calls may be the most cost-effective mobilization strategy.

Sensitivity Analysis

I conduct two kinds of sensitivity analysis. First, I investigate how the matching estimates differ between the two types of households. The instability of IV estimates for phone calls in the original data was apparent from the discrepancy between the large negative effect for single-voter households and the moderately positive effect for two-voter households. In contrast, the estimates based on matching show smaller gaps between the treatment effects for the two types of households.

I also perform one-to-one matching to examine whether it produces different estimates. One-to-one

[22] The results for the original data are as follows: 7.1% (3.2) for phone calls, 8.5% (2.1) for visits, and 2.2% (1.1) for postcards, where standard errors are in parentheses.

[23] These findings were given to me after I sent Don Green the initial version of this article.

3 Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment (Gerber, Green and Larimer, APSR)

- Setting
  - 11 Days Before 2006 Michigan Primary
  - Got rid of
    * Blocks with 10% or more of addresses being apartment numbers
    * Everyone without 9 digit zip code
    * 60% or more likely to vote in the Democratic Primary
    * Mail route with less than 25 households (postal service required at least 10 pieces of mail per route)

    - Households with multiple people with different last names

  - Randomization occurred within geographic cluster of 10,000 Households

  - Randomization was done at the household level

- Randomize 1 out of 4 letters or control:

  - 20K per Treatment Group; 100K for Control Group.

  - Civic Duty to Vote

  - You are Being Watched (Hawthorne)

  - Past Family Voting History

  - Neighbors

- Three specifications

$$
\begin{aligned}
(1.)\ Y_i &= \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \\
&\quad \beta_4 D_{4i} + u_i
\end{aligned}
$$

$$
\begin{aligned}
(2.)\ Y_i &= \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} \\
&\quad + \beta_4 D_{4i} + \gamma_k C_{ik} + u_i
\end{aligned}
$$

$$
\begin{aligned}
(3.) Y_i &= \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} \\
&\quad + \beta_4 D_{4i} + \gamma_k C_{ik} + \sum_{j=1}^{5} \lambda_j V_{ji} + u_i
\end{aligned}
$$

where Y_i is turnout of individual i in the 2006 primary, D_{ji} is a dummy for the j^{th} letter for the i^{th} person, C_{ik} is a geographical block dummy, and V_{ji} are other covariates (controls for voting in most recent 5 elections).

- Ran interactions of prior turnout and treatment with no significant effect.

TABLE 1. Relationship between Treatment Group Assignment and Covariates (Household-Level Data)

	Control	Civic Duty	Hawthorne	Self	Neighbors
	Mean	Mean	Mean	Mean	Mean
Household size	1.91	1.91	1.91	1.91	1.91
Nov 2002	.83	.84	.84	.84	.84
Nov 2000	.87	.87	.87	.86	.87
Aug 2004	.42	.42	.42	.42	.42
Aug 2002	.41	.41	.41	.41	.41
Aug 2000	.26	.27	.26	.26	.26
Female	.50	.50	.50	.50	.50
Age (in years)	51.98	51.85	51.87	51.91	52.01
N =	99,999	20,001	20,002	20,000	20,000

Note: Only registered voters who voted in November 2004 were selected for our sample. Although not included in the table, there were no significant differences between treatment group assignment and covariates measuring race and ethnicity.

that many would have decided to vote or not prior to receipt of the experimental mailings, which were sent to arrive just a few days before the election. Those considered overwhelmingly likely to favor the Democratic primary were excluded because it was thought that, given the lack of contested primaries, these citizens would tend to ignore preelection mailings. We removed everyone who lived in a route where fewer than 25 households remained, because the production process depended on using carrier-route-presort standard mail. To qualify for such treatment by the U.S. Postal Service requires that at least 10 pieces be mailed within each carrier route, which might not have been available after the control group was removed.[8] Finally, we removed all those who had abstained in the 2004 general election on the grounds that those not voting in this very high-turnout election were likely to be "deadwood"—those who had moved, died, or registered under more than one name.

Households assigned to treatment groups were sent one mailing 11 days prior to the primary election.[9] Households were randomly assigned to either the control group or one of four treatment groups described next. Each treatment group consisted of approximately 20,000 households, with 99,999 households in the control group. The 180,002 households were sorted exactly into the order required by the USPS for "ECR-LOT" eligibility (approximately: by ZIP, carrier route; then the order in which the carrier walks the route). The 180,002 households were then divided into 10,000 cells of 18 households each, with each cell consisting of households 1–18, 19–36, and so forth, of the sorted file. As a result, after sorting, each cell consisted entirely of either one or two carrier routes. A random number was generated and the entire 180,002 records were sorted by cell number and the random number. The effect was to leave all the cells together, but in a random order. Using this randomly sorted copy of the file, the records were assigned to treatments 1/1/2/2/3/3/4/4/c/c/c/c/c/c/c/c/c/c where "c" indicates "control group." The records were then resorted into carrier route order.

Table 1 shows sample statistics for subject households. The table divides the sample into treatment and control groups and shows the relationship between treatment group assignment and the covariates in the 180,002 households that form the sample for the experiment. The covariates include a set of known predictors of voting in primaries: turnout history in previous primary and general elections, gender, number of registered voters in the household, and age.

Since the randomization took place at the household level, we looked for suspicious household-level differences. Table 1 reports sample means for the households in the study and confirms that there is no relationship between a household's experimental assignment and its average level of past electoral participation. This point may be made statistically, using multinomial logit to predict experimental assignment as a function of all eight variables listed in Table 1. As expected, a likelihood ratio test with 32 degrees of freedom (8 covariates times 4 treatments) is nonsignificant ($LR = 18.6$, $p = .97$), reaffirming that the experimental groups are very closely balanced in terms of observable characteristics. Randomized assignment coupled with large sample size ensures that the unobservable characteristics are likely to be closely balanced as well.

Treatments

Each household in the treatment group received one of four mailings. The Appendix shows examples of each type. Priming voters to think about their civic duty is common to all of the treatment mailings. All four treatments carry the message "DO YOUR CIVIC DUTY—VOTE!" The first type of mailing ("Civic Duty")

Democratic nomination petitions, signing liberal initiative petitions, and living in a household with a Democrat. They were removed because of the extremely spotty pattern of contested Democratic primaries in the August 2006 election. Some people with a greater than 60% chance of voting Democratic were included, however, because they lived with another member of the household who qualified for inclusion. Such Democrats comprise 2.7% of our experimental sample. With regard to issues of external validity, we do not find any interactions between our treatments and the probability of voting Democratic.

[8] In order to achieve a universe of approximately 180,000 households, a small number of carrier routes were deleted which contained exactly 25 selected voters.

[9] These mailings are included in the Appendix.

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

provides a baseline for comparison with the other treatments because it does little besides emphasize civic duty. Households receiving this type of mailing were told, "Remember your rights and responsibilities as a citizen. Remember to vote."

The second mailing adds to this civic duty baseline a mild form of social pressure, in this case, observation by researchers. Households receiving the "Hawthorne effect" mailing were told "YOU ARE BEING STUDIED!" and informed that their voting behavior would be examined by means of public records. The degree of social pressure in this mailing was, by design, limited by the promise that the researchers would neither contact the subject nor disclose whether the subject voted. Consistent with the notion of Hawthorne effects, the purpose of this mailing was to test whether mere observation influences voter turnout.

The "Self" mailing exerts more social pressure by informing recipients that who votes is public information and listing the recent voting record of each registered voter in the household. The word "Voted" appears by names of registered voters in the household who actually voted in the 2004 primary election and the 2004 general election, and a blank space appears if they did not vote. The purpose of this mailing was to test whether people are more likely to vote if others within their own household are able to observe their voting behavior. The mailing informed voters that after the primary election "we intend to mail an updated chart," filling in whether the recipient voted in the August 2006 primary. The "Self" condition thus combines the external monitoring of the Hawthorne condition with actual disclosure of voting records.

The fourth mailing, "Neighbors," ratchets up the social pressure even further by listing not only the household's voting records but also the voting records of those living nearby. Like the "Self" mailing, the "Neighbors" mailing informed the recipient that "we intend to mail an updated chart" after the primary, showing whether members of the household voted in the primary and who among their neighbors had actually voted in the primary. The implication is that members of the household would know their neighbors' voting records, and their neighbors would know theirs. By threatening to "publicize who does and does not vote," this treatment is designed to apply maximal social pressure.

RESULTS

Following the August 2006 election we obtained turnout data from public records. Table 2 reports basic turnout rates for each of the experimental groups. The control group in our study voted at a rate of 29.7%. By comparison, the "Civic Duty" treatment group voted at a rate of 31.5%, suggesting that appeals to civic duty alone raise turnout by 1.8 percentage points. Adding social pressure in the form of Hawthorne effects raises turnout to 32.2%, which implies a 2.5 percentage-point gain over the control group. The effect of showing households their own voting records is dramatic. Turnout climbs to 34.5%, a 4.9 percentage-point increase over the control group. Even more dramatic is the effect of showing households both their own voting records and the voting records of their neighbors. Turnout in this experimental group is 37.8%, which implies a remarkable 8.1 percentage-point treatment effect.

It is important to underscore the magnitude of these effects. The 8.1 percentage-point effect is not only bigger than any mail effect gauged by a randomized experiment; it exceeds the effect of live phone calls (Arceneaux, Gerber, and Green 2006; Nickerson 2006b) and rivals the effect of face-to-face contact with canvassers conducting get-out-the-vote campaigns (Arceneaux 2005; Gerber and Green 2000; Gerber, Green, and Green 2003). Even allowing for the fact that our experiment focused on registered voters, rather than voting-eligible citizens, the effect of the Neighbors treatment is impressive. An 8.1 percentage-point increase in turnout among registered voters in a state where registered voters comprise 75% of voting-eligible citizens translates into a 6.1 percentage-point increase in the overall turnout rate. By comparison, policy interventions such as Election Day registration or vote-by-mail, which seek to raise turnout by lowering the costs of voting, are thought to have effects on the order of 3 percentage-points or less (Knack 2001).

In terms of sheer cost efficiency, mailings that exert social pressure far outstrip door-to-door canvassing. The powder blue mailings used here were printed on one side and cost 30 cents apiece to print and mail. Treating each experimental group therefore cost approximately $6,000. The "Self" mailing generated 1,854 votes at a rate of $3.24 per vote. The "Neighbors" mailing generated 3,106 votes at $1.93 per vote. By comparison, a typical door-to-door canvassing campaign produces votes at a rate of roughly $20 per vote, while phone banks tend to come in at $35 or more per vote (Green and Gerber 2004).

The analysis thus far has ignored the issue of sampling variability. The main complication associated with individual-level analysis of data that were

TABLE 3. OLS Regression Estimates of the Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Model Specifications		
	(a)	(b)	(c)
Civic Duty Treatment (Robust cluster standard errors)	.018* (.003)	.018* (.003)	.018* (.003)
Hawthorne Treatment (Robust cluster standard errors)	.026* (.003)	.026* (.003)	.025* (.003)
Self-Treatment (Robust cluster standard errors)	.049* (.003)	.049* (.003)	.048* (.003)
Neighbors Treatment (Robust cluster standard errors)	.081* (.003)	.082* (.003)	.081* (.003)
N of individuals	344,084	344,084	344,084
Covariates**	No	No	Yes
Block-level fixed effects	No	Yes	Yes

Note: Blocks refer to clusters of neighboring voters within which random assignment occurred. Robust cluster standard errors account for the clustering of individuals within household, which was the unit of random assignment.
* $p < .001$.
** Covariates are dummy variables for voting in general elections in November 2002 and 2000, primary elections in August 2004, 2002, and 2000.

randomized at the household-level is that proper estimation of the standard errors requires a correction for the possibility that individuals within each household share unobserved characteristics (Arceneaux 2005). For this reason, Table 3 reports robust cluster standard errors, which take intrahousehold correlation into account. We also consider a range of different model specifications in order to gauge the robustness of the results.

The first column of Table 3 reports the results of a linear regression in which voter turnout (Y_i) for individual i is regressed on dummy variables $\{D_{1i}, D_{2i}, D_{3i}, D_{4i}\}$ marking each of the four treatments (the reference category is the control group). This model may be written simply as

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + u_i, \quad \textbf{(6)}$$

where u_i represents an unobserved disturbance term. The second column embellishes this model by including fixed effects $\{C_{1i}, C_{2i}, \ldots, C_{9999i}\}$ for all but one of the $K = 10{,}000$ geographic clusters within which randomization occurred:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \sum_{k=1}^{K-1} \gamma_k C_{ki} + u_i. \quad \textbf{(7)}$$

The parameters associated with these fixed effects are uninteresting for our purposes; we will focus on the treatment parameters β_1, β_2, β_3, and β_4. The advantage of including fixed effects is the potential to eliminate any observed imbalances within each geographic cluster, thereby improving the precision of the estimates. The final column of Table 3 controls further for voting in five recent elections:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \sum_{k=1}^{K-1} \gamma_k C_{ki} + \lambda_1 V_{1i} + \lambda_1 V_{1i} + \cdots + \lambda_5 V_{5i} + u_i. \quad \textbf{(8)}$$

Again, the point is to minimize disturbance variance and improve the precision of the treatment estimates.

The results are remarkably robust, with scarcely any movement even in the third decimal place. The average effect of the Civic Duty mailing is a 1.8 percentage-point increase in turnout, suggesting that priming civic duty has a measurable but not large effect on turnout. The Hawthorne mailing's effect is 2.5 percentage points. Mailings that list the household's own voting record increase turnout by 4.8 percentage points, and including the voting behavior of neighbors raises the effect to 8.1 percentage points. All effects are significant at $p < .0001$. Moreover, the Hawthorne mailing is significantly more effective than the Civic Duty mailing ($p < .05$, one-tailed); the Self mailing is significantly more effective than the Hawthorne mailing ($p < .001$); and the Neighbors mailing is significantly more effective than the Self mailing ($p < .001$).

Having established that turnout increases marginally when civic duty is primed and dramatically when social pressure is applied, the remaining question is whether the effects of social pressure interact with feelings of civic duty. Using an individual's voting propensity as a proxy for the extent to which he or she feels an obligation to vote, we divided the observations into six subsamples based on the number of votes cast in five prior elections; we further divided the subsamples according to the number of voters in each household, because household size and past voting are correlated. As noted earlier, one hypothesis is that social pressure is particularly effective because it reinforces existing motivation to participate. The contrary hypothesis is that extrinsic incentives extinguish intrinsic motivation, resulting in greater treatment effects among those with low voting propensities. To test these hypotheses while at the same time taking into account floor and ceiling effects, we conducted a series of logistic regressions and examined the treatment effects across subgroups.[10] This analysis revealed that the treatment effects on underlying voting propensities are more or

[10] This analysis (not shown, but available on request) divided the subjects according to past voting history and household size. We tested the interaction hypothesis by means of a likelihood-ratio test, which failed to reject the null hypothesis of equal treatment effects across these subgroups.

APPENDIX A: MAILINGS

Civic Duty mailing

3 0 4 2 6 - 2 ||| || || || | X X X

For more information: (517) 351-1975
email: etov@grebner.com
Practical Political Consulting
P. O. Box 6249
East Lansing, MI 48826

PRSRT STD
U.S. Postage
PAID
Lansing, MI
Permit # 444

ECRLOT **C002
THE JONES FAMILY
9999 WILLIAMS RD
FLINT MI 48507

Dear Registered Voter:

DO YOUR CIVIC DUTY AND VOTE!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

The whole point of democracy is that citizens are active participants in government; that we have a voice in government. Your voice starts with your vote. On August 8, remember your rights and responsibilities as a citizen. Remember to vote.

DO YOUR CIVIC DUTY — VOTE!

Hawthorne mailing

3 0 4 2 4 - 1

For more information: (517) 351-1975
email: etov@grebner.com
Practical Political Consulting
P. O. Box 6249
East Lansing, MI 48826

PRSRT STD
U.S. Postage
PAID
Lansing, MI
Permit # 444

ECRLOT **C001
THE SMITH FAMILY
9999 PARK LANE
FLINT MI 48507

Dear Registered Voter:

YOU ARE BEING STUDIED!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

This year, we're trying to figure out why people do or do not vote. We'll be studying voter turnout in the August 8 primary election.

Our analysis will be based on public records, so you will not be contacted again or disturbed in any way. Anything we learn about your voting or not voting will remain confidential and will not be disclosed to anyone else.

DO YOUR CIVIC DUTY — VOTE!

Self mailing

3 0 4 2 2 - 4

For more information: (517) 351-1975
email: etov@grebner.com
Practical Political Consulting
P. O. Box 6249
East Lansing, MI 48826

PRSRT STD
U.S. Postage
PAID
Lansing, MI
Permit # 444

ECRLOT **C050
THE WAYNE FAMILY
9999 OAK ST
FLINT MI 48507

Dear Registered Voter:

WHO VOTES IS PUBLIC INFORMATION!

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse.

This year, we're taking a different approach. We are reminding people that who votes is a matter of public record.

The chart shows your name from the list of registered voters, showing past votes, as well as an empty box which we will fill in to show whether you vote in the August 8 primary election. We intend to mail you an updated chart when we have that information.

We will leave the box blank if you do not vote.

DO YOUR CIVIC DUTY—VOTE!

OAK ST	Aug 04	Nov 04	Aug 06
9999 ROBERT WAYNE		Voted	______
9999 LAURA WAYNE	Voted	Voted	______

Neighbors mailing

3 0 4 2 3 - 3

For more information: (517) 351-1975
email: etov@grebner.com
Practical Political Consulting
P. O. Box 6249
East Lansing, MI 48826

PRSRT STD
U.S. Postage
PAID
Lansing, MI
Permit # 444

ECRLOT **C050
THE JACKSON FAMILY
9999 MAPLE DR
FLINT MI 48507

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR		Aug 04	Nov 04	Aug 06
9995	JOSEPH JAMES SMITH	Voted	Voted	______
9995	JENNIFER KAY SMITH		Voted	______
9997	RICHARD B JACKSON		Voted	______
9999	KATHY MARIE JACKSON		Voted	______
9999	BRIAN JOSEPH JACKSON		Voted	______
9991	JENNIFER KAY THOMPSON		Voted	______
9991	BOB R THOMPSON		Voted	______
9993	BILL S SMITH			______
9989	WILLIAM LUKE CASPER		Voted	______
9989	JENNIFER SUE CASPER		Voted	______
9987	MARIA S JOHNSON	Voted	Voted	______
9987	TOM JACK JOHNSON	Voted	Voted	______
9987	RICHARD TOM JOHNSON		Voted	______
9985	ROSEMARY S SUE		Voted	______
9985	KATHRYN L SUE		Voted	______
9985	HOWARD BEN SUE		Voted	______
9983	NATHAN CHAD BERG		Voted	______
9983	CARRIE ANN BERG		Voted	______
9981	EARL JOEL SMITH			______
9979	DEBORAH KAY WAYNE		Voted	______
9979	JOEL R WAYNE		Voted	______

1 Empathy or Antipathy: The Impact of Diversity

- Setting
  - UCLA gives questionnaire to determine residence for entering freshmen.
  - Randomized after allocating based upon:
    * smoking/nonsmoking room
    * substance-free housing
    * single/double/triple occupancy
    * geographic area of campus
    * gender composition of corridor
  - In follow up surveys (mostly after graduation), got 89-90% response rate for those entering 1997-1999

– Look at impact of initial assignment of roommate's race on white student preferences towards affirmative action

TABLE 1—SAMPLE ATTRITION

	Total	1997	1998	1999	2000
Response rate on CIRP survey for all entering students		89%	89%	90%	n/a
Number of students responding to CIRP survey of which:	14,235	3,967	3,573	3,419	3,276
Students opting to live in enrichment dormitories	3,246	1,014	920	633	679
Students requesting a specific roommate	2,354	325	755	662	612
Students failing to meet the lottery deadline	5,583	1,449	1,166	1,615	1,353
Students living alone during the first year	979	255	273	215	236
Students not assigned roommates	63	21	5	12	25
Total number of students randomly assigned roommates of which:	2,010	903	454	282	371
Students designated race as "black" only	47	19	8	8	12
Students designated race as "white" only	1,647	729	377	236	305
Students designated race as "Hispanic" (see text)	61	26	14	7	14
Students designated race as "Asian" (see text)	149	72	34	19	24
Students with other racial designations	106	57	21	12	16
Target sample of white students opting for random assignment of which:	1,647	729	377	236	305
Failed to respond to follow-up survey	369	133	91	75	70
Response rate on follow-up survey	78%	82%	76%	68%	77%
Final analysis sample	1,278	596	286	161	235

designated themselves as "white." The follow-up survey response rate among this sample was 78 percent and produced an analysis sample of 1,278. Missing data on individual survey items reduced this case count further. We address the issue of possible nonresponse bias below.

Outcome measures were derived from sections in the follow-up survey corresponding to three broad domains: attitudes, behaviors, and goals. Questions on racial attitudes in the survey ask for strong agreement (coded as 4), agreement (3), disagreement (2), or strong disagreement (1) with the following statements: (a) "Affirmative action in college admission should be abolished"; (b) "Affirmative action is justified if it ensures a diverse student body on college campuses"; and (c) "Having a diverse student body is essential for high quality education."[4] The first of these items was also asked with identical wording on the 1997, 1999, and 2000 entering-student CIRP survey. Neither the second nor third items was asked in any of the CIRP surveys.

On the behavior front, respondents to our follow-up survey were also asked to specify the number of times per month when "I have personal contact with people from other racial/ethnic groups"; when "I interact comfortably with people from other racial/ethnic groups"; and when "I socialize with someone with an African American background."

The section on goals in both the CIRP and the follow-up survey contained questions about major life goals such as "becoming an authority in my field" and "being very well off financially." In terms of goals related to race, respondents were asked how imperative the following goals were to them personally: "helping promote racial understanding"; "helping others who are in difficulty"; "working to eliminate discrimination against people of color"; and "participating actively in civil rights organizations." All goals were rated on a scale of essential (coded as 4), very important (3), important (2), and not important (1).

Given the ordinal nature of the key attitudinal outcomes, we used ordered probit regression. Results from comparable OLS models, which presume a cardinal scale for the attitudinal responses but also increase the precision of the estimates, are shown in our tables for purposes of comparison. In all cases, responses were scaled so the higher scores indicated more "liberal" attitudes and behaviors. Since a number of these and related questions were included in the entering-student CIRP survey, we include baseline controls for the respondent's own responses (standardized and scaled in a "liberal" direction) to the following statements: (a) "Affirmative action in college admissions should be abolished"; (b) "Race discrimination is no longer a major problem in America"; and (c) "Colleges should prohibit racist/sexist speech on campus." To control for class-related

[4] We explored with factor analysis whether these or any other attitudinal items could be combined into an index, but in no case were the correlations among three items high enough to warrant this.

TABLE 2—MEANS AND STANDARD DEVIATIONS OF RESPONDENTS' AND NONRESPONDENTS' CHARACTERISTICS FROM THE ENTERING STUDENT SURVEYS

	All respondents to the follow-up survey (1)	White respondents to the follow-up survey (all randomly-assigned roommates) (2)	White respondents to CIRP entering survey but not randomly-assigned roommates (3)	White randomly assigned roommates who FAILED to respond to the follow-up survey (4)	*p* value of *t*-test or Chi-square test comparing (4) and (2) (5)	Blacks randomly-assigned roommates (6)	*p* value of *t*-test or Chi-square test comparing (6) and (2) (7)	Black respondents to CIRP Entering Survey but not randomly-assigned roommates (8)	*p* value of *t*-test or Chi-square test comparing (8) and (6) (9)
Affirmative action in college admissions should be abolished (reversed)[a]	2.083	2.016	2.033	2.089	0.110	3.487	0.000	3.240	0.020
	(0.813)	(0.772)	(0.774)	(0.763)		(0.547)		(0.714)	
Race discrimination is no longer a major problem in America (reversed)[a]	3.215	3.166	3.172	3.238	0.093	3.558	0.000	3.650	0.323
	(0.719)	(0.723)	(0.730)	(0.733)		(0.682)		(0.615)	
Colleges should prohibit racist/sexist speech on campus[b]	2.477	2.434	2.424	2.504	0.211	2.617	0.196	2.853	0.120
	(0.956)	(0.942)	(0.958)	(0.952)		(1.153)		(1.003)	
Wealthy people should pay a larger share of taxes than they do now[b]	2.524	2.518	2.489	2.426	0.086	2.620	0.463	2.743	0.366
	(0.927)	(0.928)	(0.929)	(0.860)		(1.089)		(0.891)	
Father's education	16.360	16.362	16.399	16.565	0.070	15.834	0.066	15.089	0.033
	(1.980)	(1.921)	(1.977)	(1.831)		(2.230)		(2.334)	
Mother's education	15.801	15.810	15.911	15.903	0.433	15.957	0.623	15.155	0.014
	(2.083)	(2.023)	(1.978)	(1.946)		(1.922)		(2.185)	
High-school grade point average	3.762	3.775	3.752	3.741	0.023	3.543	0.000	3.480	0.324
	(0.260)	(0.251)	(0.280)	(0.276)		(0.366)		(0.423)	
Test scores (ACT scale)	28.051	28.209	28.372	27.888	0.034	25.134	0.000	24.118	0.060
	(2.616)	(2.594)	(2.854)	(2.457)		(2.952)		(3.630)	
Family income < $50,000	0.114	0.105	0.112	0.068		0.170		0.392	
Family income $50,000 to $74,999	0.166	0.159	0.150	0.138		0.213		0.200	
Family income $75,000 to $149,999	0.405	0.417	0.375	0.388	0.001	0.340	0.547	0.255	0.000
Family income $150,000 to $199,999	0.094	0.101	0.098	0.098		0.128		0.038	
Family income ≥ $200,000	0.121	0.119	0.167	0.198		0.128		0.032	
Missing Family Income	0.100	0.099	0.098	0.111		0.021		0.083	
	n = 1,558	n = 1,278	n = 9,099	n = 369		n = 47		n = 832	

Note: Blacks randomly assigned roommates may or may not have been respondents to the follow-up survey.

[a] Scale: (4) disagree strongly; (3) disagree somewhat; (2) agree somewhat; (1) agree strongly.

[b] Scale: (4) agree strongly; (3) agree somewhat; (2) disagree somewhat; (1) disagree strongly.

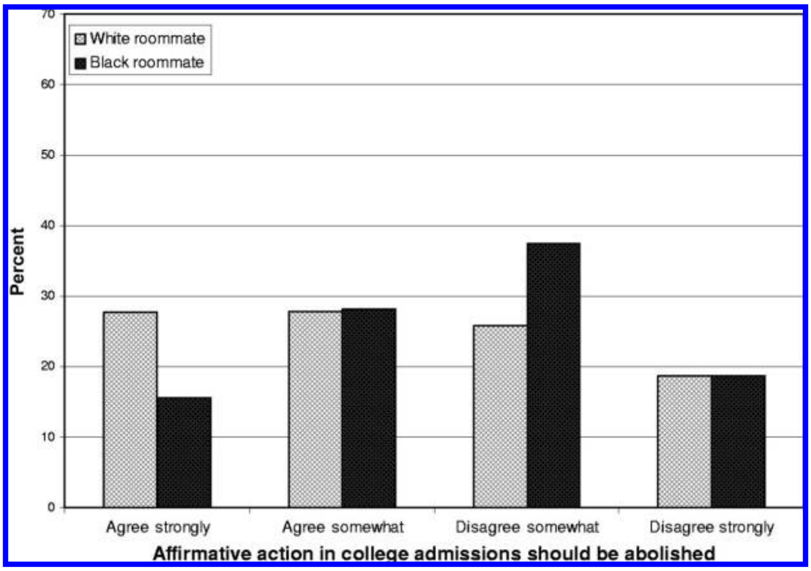


FIGURE 1. ROOMMATE RACE AND ATTITUDES TOWARD AFFIRMATIVE ACTION

questions was between one-third and one-half of a standard deviation higher among whites who were randomly assigned black roommates than among whites assigned white roommates. Estimated effects on endorsement of the proposition that "a diverse student body is essential for high-quality education" exceed half a standard deviation in the ordered probit regressions. The estimated effect sizes translate into increments in the four-point, agree-disagree scale of one-third to three-quarters of a point. Responses to these attitudinal questions for white students assigned other minority roommates did not differ significantly from white students assigned white roommates.[7]

Not surprisingly, the respondents' prior responses to affirmative action and income redistribution questions in the entering-student CIRP questionnaire were strong significant predictors of affirmative action responses 1.5 to 6.5 years later in several cases (results available upon request). The respondent's own SAT/ACT test scores had an inconsistently negative impact on current affirmative action attitudes, while maternal schooling had an inconsistently positive association with them.

Students who were assigned black roommates during their first year report more frequent personal contact and comfortable interactions with members of other racial/ethnic groups in later years (Table 4, columns 1 and 2). But while reported contact and comfort with minorities increased, reported friendships and socializing did not change significantly (Table 4, columns 3 and 4).[8] In no instance was assignment to other minority roommates a significant predictor of these four outcomes.

The follow-up survey also asked respondents how long they had lived with their roommates; how often they socialized with their initial roommates both during the first year and in the twelve months prior to the follow-up survey; and how friendly they still were with their initial roommates. Since these questions were not asked for each specific randomly assigned roommate, we restricted the sample of white students from the 1,278 who responded to the follow-up survey to the

[7] When we broke the "other minority" category into "Asian," "Hispanic," and "mixed," we found no significant differences between any of these categories and the omitted, white roommate, category.

[8] While we were able to control for baseline measures of the outcome in the regressions where the dependent variable was an attitude, we were not able to do so in the regressions where the dependent variable was a behavior (because we lacked baseline data on behaviors). Other things being equal, this makes it harder to detect a statistically significant roommate effect in the behavior regressions than in the attitudinal ones.

TABLE 3—ORDERED PROBIT AND OLS REGRESSIONS COEFFICIENTS, AND STANDARD ERRORS FOR ROOMMATE PREDICTORS OF ATTITUDES OF WHITE STUDENTS TWO TO SIX YEARS AFTER ENTERING COLLEGE

	Affirmative action in college admissions should be abolished (reverse coding)[a]			Affirmative action is justified if it ensures a diverse student body on college campuses[b]			Having a diverse student body is essential for high-quality education[b]		
	Ordered probit regressions		OLS regression	Ordered probit regressions		OLS regression	Ordered probit regression		OLS regression
ROOMMATES' CHARACTERISTICS									
Any black roommate(s)	0.497**	0.489**	0.366*	0.493**	0.506**	0.429**	0.743***	0.770***	0.470***
	(0.239)	(0.249)	(0.219)	(0.236)	(0.239)	(0.206)	(0.256)	(0.293)	(0.154)
Any other minority roommate(s)	0.027	0.029	0.032	0.096	0.154	0.120	0.022	0.056	0.025
	(0.099)	(0.107)	(0.096)	(0.100)	(0.107)	(0.094)	(0.104)	(0.106)	(0.072)
Only white roommate(s) [omitted group]	—	—	—	—	—	—	—	—	—
At least one roommate with family income < $50,000		0.125	0.105		−0.012	−0.006		0.319**	0.180*
		(0.129)	(0.113)		(0.131)	(0.112)		(0.136)	(0.087)
At least one roommate with family income between $50,000 and $74,999		0.043	0.016		0.055	0.032		0.055	0.047
		(0.108)	(0.096)		(0.107)	(0.093)		(0.112)	(0.075)
At least one roommate with family income between $75,000 and $149,999 [omitted group]		—	—		—	—		—	—
At least one roommate with family income between $150,000 and $199,999		0.078	0.061		0.069	0.056		0.061	0.043
		(0.130)	(0.115)		(0.129)	(0.109)		(0.131)	(0.087)
At least one roommate with family income ≥ $200,000		−0.023	−0.020		−0.077	−0.061		0.156	0.097
		(0.112)	(0.100)		(0.115)	(0.100)		(0.123)	(0.082)
TIME									
Years since sophomore year		0.165	0.130		0.095	0.082		−0.104	−0.077
		(0.113)	(0.098)		(0.108)	(0.092)		(0.109)	(0.081)
R-squared/Pseudo-R^2		0.180	0.370		0.178	0.371		0.191	0.356
Number of observations	1,172	1,169	1,169	1,196	1,193	1,193		1,241	1,241

Notes: Standard errors are given in parentheses. Standard errors are adjusted for room clustering using Huber-White robust estimations. All regressions include controls for respondent's: father's education, mother's education, family income, high-school grade point average, ACT/SAT score, CIRP-based attitudes about race discrimination, taxation of the rich and prohibition of racist/sexist speech. For roommates': average father's education, average mother's education, average high school grade-point average, average ACT/SAT score. All regressions also control for respondent housing preferences, gender, cohort, test taken; values not shown. "—" indicates that the variable was not included in the regression.

[a] Scale: (4) disagree strongly; (3) disagree somewhat; (2) agree somewhat; (1) agree strongly.

[b] Scale: (4) agree strongly; (3) agree somewhat; (2) disagree somewhat; (1) disagree strongly.

* $p <= 0.10$. ** $p <= 0.05$. *** $p <= 0.01$.

TABLE 4—OLS REGRESSION COEFFICIENTS AND STANDARD ERRORS FOR ROOMMATE PREDICTORS OF BEHAVIORS OF WHITE STUDENTS TWO TO SIX YEARS AFTER ENTERING COLLEGE

	I have personal contact with people from other racial/ethnic groups (*number of times per month*)	I interact comfortably with people from other racial/ethnic groups (*number of times per month*)	Fraction of friends from own racial/ethnic background	Socialized with someone with an African American background (*number of times per month*)
ROOMMATES' CHARACTERISTICS				
Any black roommate(s)	2.949*	2.844**	−0.048	1.830
	(1.730)	(1.436)	(0.045)	(1.826)
Any other minority roommate(s)	0.052	0.214	−0.011	−0.982
	(0.794)	(0.740)	(0.016)	(0.911)
Only white roommate(s) [omitted group]	—	—	—	—
At least one roommate with family income < \$50,000	0.719	1.042	0.026	2.306**
	(0.963)	(0.895)	(0.019)	(1.073)
At least one roommate with family income between \$50,000 and \$74,999	0.996	0.267	−0.024	1.7622*
	(0.754)	(0.744)	(0.018)	(0.968)
At least one roommate with family income between \$75,000 and \$149,999 [omitted group]	—	—	—	—
At least one roommate with family income between \$150,000 and \$199,999	0.851	0.883	−0.010	1.382
	(0.918)	(0.871)	(0.019)	(1.127)
At least one roommate with family income ≥ \$200,000	0.592	1.349*	−0.007	1.064
	(0.868)	(0.741)	(0.019)	(1.026)
TIME				
Years since sophomore year	−0.743	−0.689	0.006	−1.333
	(0.820)	(0.802)	(0.015)	(0.918)
R-squared	0.189	0.201	0.171	0.230
Number of observations	1,257	1,254	1,245	1,243

Notes: Standard errors are given in parentheses. Standard errors are adjusted for room clustering using Huber-White robust estimations. All regressions include controls for respondents: father's education, mother's education, family income, high-school grade-point average, ACT/SAT score, CIRP-based attitudes about race discrimination, taxation of the rich, and prohibition of racist/sexist speech. For roommates: average father's education, average mother's education, average high-school grade-point average, average ACT/SAT score. All regressions also control for respondent housing preferences, gender, cohort, test taken; values not shown. "—" indicates that the variable was not included in the regression.

* $p <= 0.10$. ** $p <= 0.05$. *** $p <= 0.01$.

1,087 white students who had only one roommate. The vast majority (923, or 85 percent) had white roommates; 21 had black roommates, 70 had Asian roommates, 25 had Hispanic roommates, and 48 had "other" race roommates. We found no statistically significant differences in frequency of subsequent interactions depending on roommate race. For example, 14 percent of whites with white roommates and 15 percent of whites with black roommates considered these roommates to be their "best college friend." Very close fractions (41 percent and 45 percent, respectively) were either "not in touch" or "did not get along" with these roommates. Similar fractions (14 percent and 10 percent) had socialized more than once a week with their first-year roommates in the past year, while 62 percent and 50 percent had socialized more than once a week with their initial roommates during their first year. Keeping in mind the low power for this analysis, there did not appear to be appreciable differences in the duration or nature of friendships white students struck with white and black roommates.

B. *Extensions*

We explored several extensions of the analysis above. First, we investigated whether the effects of being assigned a black roommate persisted over time. Second, we explored

2 Women as Policy Makers: Evidence from A Randomized Policy Experiment in India

- To encourage female political leadership, India mandated that each state had to have 1/3 of seats in Panchayat (Village) Councils as well as 1/3 of Pradhan (Mayor) female.

- Already seats had been designated for scheduled castes and tribes (See Pande, AER).

- As of 2004, all states except Bihar and Uttar Pradesh had implemented the mandate.

- To implement the central mandate, in 1993, West Bengal reserved 1/3 of the councilor positions but only 196 out of 3,324 became Pradhan; in 1998, the law was changed to require both 1/3 of councilor position and 1/3 of pradhans to be female.

- In Rajasthan, 1/3 were required to be female at both levels as of 1995 (and thus in 2000).

- Paper collected data in two districts:

  - Birbhum, West Bengal (125 miles from Calcutta)

  - Udaipur, Rajasthan

- Started in summer of 2000 in West Bengal: surveyed both Pradhan and villagers

  - 166 villages; 5 used in pilot, 161 others were surveyed.

- Udaipur was surveyed between August 2002 and December 2002

  - 100 villages chosen at random

- Estimation

$$Y_{ij} = \beta_1 + \beta_2 R_j + \beta_3 D_i * R_j + \sum_{l=1}^{N} \beta_l d_{il} + \epsilon_{ij} \quad (1)$$

$$Y_{ij} = \beta_4 + \beta_5 R_j + \beta_6 S_i * R_j + \sum_{l=1}^{N} \beta_l d_{il} + \epsilon_{ij} \quad (2)$$

$$\begin{aligned} Y_{ij} \;=\; & \beta_7 + \beta_8 R_j + \beta_9 D_i * R_j + \beta_{10} D_{ij} R_j + \qquad (3) \\ & \beta_{11} S_{ij} * R_j + \beta_{12} S_i + \beta_{13} D_{ij} + \sum_{l=1}^{N} \beta_l d_{il} + \epsilon_{ij} \end{aligned}$$

where R_j is a dummy for reservation for a woman, D_i is a dummy for is the difference in fractions of requests about good i from women, S_i is the average fraction of requests across men and women, D_{ij} is the difference between an indicator for whether issue i was brought by women in village j or by men in village j, and S_{ij} is the sum of an indicator for whether issue i was brought by women in village j and men in village j.

number (an administrative number pre-dating this reform). They are then ranked in three separate lists, according to whether or not the seats were reserved for a SC, for a ST, or were unreserved (these reservations were also chosen randomly, following a similar method). Using these lists, every third GP starting with the first on the list is reserved for a woman Pradhan for the first election.[6]

From discussions with the government officials at the Panchayat Directorate who devised the system and district officials who implemented it in individual districts, it appears that these instructions were successfully implemented. More importantly, in the district we study in West Bengal, we could verify that the policy was strictly implemented. After sorting the GPs into those reserved for SC/ST and those not reserved, we could reconstruct the entire list of GPs reserved for a woman by sorting all GPs by their serial number, and selecting every third GP starting from the first in each list. This verifies that the allocation of GPs to the reserved list was indeed random, as intended.[7]

Table I shows the number of female Pradhans in reserved and unreserved GPs in both states. In both states, all Pradhans in GPs reserved for a woman are female. In West Bengal, only 6.5% of the Pradhans are female in unreserved GPs. In Rajasthan, only one woman was elected on an unreserved seat, despite the fact that this was the second cycle. Women elected once due to the reservation system were not re-elected.[8]

TABLE I

FRACTION OF WOMEN AMONG PRADHANS IN RESERVED AND UNRESERVED GP

	Reserved GP (1)	Unreserved GP (2)
West Bengal		
Total Number	54	107
Proportion of Female Pradhans	100%	6.5%
Rajasthan		
Total Number	40	60
Proportion of Female Pradhans	100%	1.7%

[6]For the next election, every third GP starting with the second on the list was reserved for a woman, etc. The Panchayat Constitution Rule has actual tables indicating the ranks of the GPs to be reserved in each election.

[7]We could not obtain the necessary information to perform the same exercise in Rajasthan. However, there too, the system appears to have been correctly implemented.

[8]The one woman elected on an unreserved seat had not been previously elected on a reserved seat.

TABLE II

VILLAGE CHARACTERISTICS IN RESERVED AND UNSERVED GP, 1991 CENSUS

	West Bengal			Rajasthan		
Dependent Variables	Mean, Reserved GP (1)	Mean, Unreserved GP (2)	Difference (3)	Mean, Reserved GP (4)	Mean, Unreserved GP (5)	Difference (6)
Total Population	974	1022	−49	1249	1564	−315
	(60)	(46)	(75)	(123)	(157)	(212)
Female Literacy Rate	.35	.34	.01	.05	.05	.00
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Male Literacy Rate	.57	.58	−.01	.28	.26	.03
	(.01)	(.01)	(.01)	(.02)	(.02)	(.03)
% Cultivated Land that Is Irrigated	.45	.43	.02	.05	.07	−.02
	(.03)	(.02)	(.04)	(.01)	(.01)	(.02)
Dirt Road	.92	.91	.01	.40	.52	−.11
	(.02)	(.01)	(.02)	(.08)	(.07)	(.10)
Metal Road	.18	.15	.03	.31	.34	−.04
	(.03)	(.02)	(.03)	(.07)	(.06)	(.10)
Bus Stop or Train Station	.31	.26	.05	.40	.43	−.03
	(.04)	(.02)	(.04)	(.08)	(.07)	(.10)
Number of Public Health Facilities	.06	.08	−.02	.29	.19	.10
	(.01)	(.01)	(.02)	(.08)	(.06)	(.10)
Tube Well Is Available	.05	.07	−.02	.02	.03	−.01
	(.03)	(.02)	(.07)	(.02)	(.02)	(.03)
Handpump Is Available	.84	.88	−.04	.90	.97	−.06
	(.04)	(.03)	(.05)	(.05)	(.02)	(.05)
Wells	.44	.47	−.02	.93	.91	.01
	(.07)	(.04)	(.08)	(.04)	(.04)	(.06)
Tap Water	.05	.03	.01	.12	.09	.03
	(.03)	(.02)	(.03)	(.05)	(.04)	(.06)
Number of Primary Schools	.95	.91	.04	.93	1.16	−.23
	(.07)	(.03)	(.08)	(.09)	(.10)	(.15)
Number of Middle Schools	.05	.05	.00	.43	.33	.10
	(.01)	(.01)	(.01)	(.08)	(.07)	(.10)
Number of High Schools	.09	.10	−.01	.14	.07	.07
	(.01)	(.01)	(.02)	(.06)	(.04)	(.07)
F-Statistics: Difference Jointly Significant			.93			1.54
(*p*-value)			(.53)			(.11)

Notes: 1. There are 2120 observations in the West Bengal regressions, and 100 in the Rajasthan regressions. 2. Standard errors, corrected for clustering at the GP level in the West Bengal regressions, are in parentheses.

TABLE III

EFFECT OF WOMEN'S RESERVATION ON WOMEN'S POLITICAL PARTICIPATION

Dependent Variables	Mean, Reserved GP (1)	Mean, Unreserved GP (2)	Difference (3)
West Bengal			
Fraction of Women Among Participants in the Gram Samsad (in percentage)	9.80	6.88	2.92
	(1.33)	(.79)	(1.44)
Have Women Filed a Complaint to the GP in the Last 6 Months	.20	.11	.09
	(.04)	(.03)	(.05)
Have Men Filed a Complaint to the GP in the Last 6 Months	.94	1.00	.06
	(.06)		(.06)
Observations	54	107	
Rajasthan			
Fraction of Women Among Participants in the Gram Samsad (in percentage)	20.41	24.49	−4.08
	(2.42)	(3.05)	(4.03)
Have Women Filed a Complaint to the GP in the Last 6 Months	.64	.62	.02
	(.07)	(.06)	(.10)
Have Men Filed a Complaint to the GP in the Last 6 Months	.95	.88	.073
	(.03)	(.04)	(.058)
Observations	40	60	

Notes: 1. Standard errors in parentheses. 2. Standard errors are corrected for clustering at the GP level in the West Bengal regressions, using the Moulton (1986) formula.

percentage of eligible voters attending the Gram Samsad, this corresponds to a net increase in the participation of women, and a decline in the participation of men. This is consistent with the idea that political communication is influenced by the fact that citizens and leaders are of the same sex. Women in villages with a reserved Pradhan are twice as likely to have addressed a request or a complaint to the GP Pradhan in the last 6 months, and this difference is significant.[22] The fact that the Pradhan is a woman therefore significantly increases the involvement of women in the affairs of the GP in West Bengal.

In Rajasthan, the fact that the Pradhan is a woman has no effect on women's participation at the Gram Samsad or the occurrence of women's complaints. Note that women participate more in the Gram Samsad in Rajasthan, most probably because the process is very recent, and the GP leaders are trained to mobilize women in public meetings.[23]

[22] In the subsample of villages in which we conducted follow-up surveys, we also asked whether men had brought up any issue in the previous six months. In all cases but one (a reserved GP), they had.

[23] Interestingly, women's participation is significantly higher when the position of council member *of the village* is reserved for a woman (results not reported to conserve space). This difference is probably due to the very long distance between villages in Rajasthan.

TABLE IV

ISSUES RAISED BY WOMEN AND MEN IN THE LAST 6 MONTH

	West Bengal						Rajasthan					
	Women			Men	Average	Difference	Women			Men	Average	Difference
	Reserved	Unreserved	All				Reserved	Unreserved	All			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Other Programs												
Public Works	.84	.84	.84	.85	.84	−.01	.60	.64	.62	.87	.74	−.26
Welfare Programs	.12	.09	.10	.04	.07	.06	.25	.14	.19	.03	.04	.16
Child Care	.00	.02	.01	.01	.01	.00	.04	.09	.07	.01	.02	.06
Health	.03	.04	.04	.02	.03	.02	.06	.08	.07	.04	.03	.03
Credit or Employment	.01	.01	.01	.09	.05	−.08	.06	.06	.05	.04	.09	.01
Total Number of Issues	153	246	399	195			72	88	160	155		
Breakdown of Public Works Issues												
Drinking Water	.30	.31	.31	.17	.24	.13	.63	.48	.54	.43	.49	.09
Road Improvement	.30	.32	.31	.25	.28	.06	.09	.14	.13	.23	.18	−.11
Housing	.10	.11	.11	.05	.08	.05	.02	.04	.03	.04	.04	−.01
Electricity	.11	.07	.08	.10	.09	−.01	.02	.04	.03	.02	.02	.01
Irrigation and Ponds	.02	.04	.04	.20	.12	−.17	.02	.02	.02	.04	.03	−.02
Education	.07	.05	.06	.12	.09	−.06	.02	.07	.05	.13	.09	−.09
Adult Education	.01	.00	.00	.01	.00	.00	0	0	.00	.00	.00	.00
Other	.09	.11	.10	.09	.09	.01	.19	.21	.20	.12	.28	.05
Number of Public Works Issues	128	206	334	166			43	56	99	135		
Public Works												
Chi-square	8.84		71.72				7.48		16.38			
p-value	.64		.00				.68		.09			

Notes: 1. Each cell lists the number of times an issue was mentioned, divided by the total number of issues in each panel. 2. The data for men in West Bengal comes from a subsample of 48 villages. 3. Chi-square values placed across two columns test the hypothesis that issues come from the same distribution in the two columns.

TABLE V

EFFECT OF WOMEN'S RESERVATION ON PUBLIC GOODS INVESTMENTS

	West Bengal			Rajasthan		
	Mean, Reserved GP	Mean, Unreserved GP	Difference	Mean, Reserved GP	Mean, Unreserved GP	Difference
Dependent Variables	(1)	(2)	(3)	(4)	(5)	(6)
A. *Village Level*						
Number of Drinking Water Facilities	23.83	14.74	9.09	7.31	4.69	2.62
Newly Built or Repaired	(5.00)	(1.44)	(4.02)	(.93)	(.44)	(.95)
Condition of Roads (1 if in good	.41	.23	.18	.90	.98	−.08
condition)	(.05)	(.03)	(.06)	(.05)	(.02)	(.04)
Number of Panchayat Run	.06	.12	−.06			
Education Centers	(.02)	(.03)	(.04)			
Number of Irrigation Facilities	3.01	3.39	−.38	.88	.90	−.02
Newly Built or Repaired	(.79)	(.8)	(1.26)	(.05)	(.04)	(.06)
Other Public Goods (ponds, biogas,	1.66	1.34	.32	.19	.14	.05
sanitation, community buildings)	(.49)	(.23)	(.48)	(.07)	(.06)	(.09)
Test Statistics: Difference Jointly Significant			4.15			2.88
(*p*-value)			(.001)			(.02)
B. *GP Level*						
1 if a New Tubewell Was Built	1.00	.93	.07			
		(.02)	(.03)			
1 if a Metal Road Was Built or Repaired	.67	.48	.19			
	(.06)	(.05)	(.08)			
1 if There Is an Informal Education	.67	.82	−.16			
Center in the GP	(.06)	(.04)	(.07)			
1 if at Least One Irrigation Pump Was Built	.17	.09	.07			
	(.05)	(.03)	(.05)			
Test Statistics: Difference Jointly Significant			4.73			
(*p*-value)			(.001)			

Notes: 1. Standard errors in parentheses. 2. In West Bengal, there are 322 observations in the village level regressions, and 161 in the GP level regressions. There are 100 observations in the Rajasthan regressions. 3. Standard errors are corrected for clustering at the GP level in the village level regressions, using the Moulton (1986) formula, for the West Bengal regressions.

TABLE VI

OLS REGRESSIONS: DETERMINANTS OF PUBLIC GOOD PROVISION

	West Bengal					Rajasthan			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Reserved for a Woman	.23	−.17	.00	.18	.17	.16	−.29	.04	.16
	(.101)	(.123)	(.159)	(.136)	(.111)	(.115)	(.19)	(.16)	(.118)
Reserved $* D_i$	1.63		1.22	1.56	1.67	4.40		4.66	4.29
	(.501)		(.799)	(.629)	(.554)	(1.454)		(1.6)	(1.491)
Reserved $* S_i$		2.04					1.78		
		(.642)					(.728)		
Reserved $* D_{\{ij\}}$			.03					−.37	
(village level)			(.047)					(.169)	
Reserved $* S_{\{ij\}}$			−.01					.05	
(village level)			(.155)					(.27)	
Pradhan is New					−.09				
					(.079)				
Pradhan is New $* D_i$					−.10				
					(.323)				
Reservation in 2003					.03				
					(.093)				
Reservation in 2003 $* D_i$					−.19				
					(.326)				
Reserved for SC/ST					−.07				.00
					(.075)				(.18)
Reserved for SC/ST $* D_i$					.10				.03
					(.145)				(.315)
$D_{\{ij\}}$	No	No	Yes	No	No	No	No	Yes	No
$S_{\{ij\}}$	No	No	Yes	No	No	No	No	Yes	No
Pradhan's Characteristics	No	No	No	Yes	No	No	No	No	No
Pradhan's Characteristics $* D_i$	No	No	No	Yes	No	No	No	No	No

Notes: 1. The dependent variable is a standardized measure of investment in each good. There are six types of goods in West Bengal (drinking water, roads, informal education, formal education, irrigation, others) and four types of goods in Rajasthan (drinking water, roads, formal education, others). 2. Standard errors (corrected for clustering at the GP level using Moulton (1986) in West Bengal) are in parentheses below the coefficients. 3. The regressions include a good-specific fixed effect. 4. The variables D_i, S_i, $D_{\{ij\}}$, and $S_{\{ij\}}$ are defined in the text: D_i is the relative strength of women's preference for good i in the district; S_i is the average strength of preference in the district; $D_{\{ij\}}$ is the difference of indicators for whether good i was mentioned by women and men in village j; $S_{\{ij\}}$ is the sum of the indicators for whether good i was mentioned by women and men in village j. 5. Pradhan characteristics include all variables in Table VII. 6. There are 323 village level observations in West Bengal, and 100 village level observations in Rajasthan.

TABLE VII

PRADHAN'S CHARACTERISTICS IN RESERVED AND UNSERVED GP (WEST BENGAL)

	West Bengal		
	Mean, Reserved GP	Mean, Unreserved GP	Difference
Dependent Variables	(1)	(2)	(3)
A. *Pradhan's Background*			
Age	31.87	39.72	−7.85
	(1.08)	(.87)	(1.45)
Years of Education	7.13	9.92	−2.79
	(.48)	(.29)	(.54)
Literacy	.80	.98	−.19
	(.06)	(.01)	(.04)
Married	.89	.87	.02
	(.04)	(.03)	(.06)
Number of Children	2.45	2.50	−.05
	(.20)	(.15)	(.26)
Below Poverty Line	.46	.28	.18
	(.07)	(.04)	(.08)
Number of Household Assets	1.72	2.36	−.64
	(.18)	(.14)	(.23)
Population of Pradhan's Own Village	1554	2108	−554
	(204)	(179)	(291)
Hesitates when Answering the Questions	.75	.41	.34
(interviewer's impression)	(.06)	(.05)	(.08)
B. *Pradhan's Political Aspirations and Experience*			
Was Elected to the GP Council Before 1998	.11	.43	−.32
	(.04)	(.05)	(.07)
Was Elected Pradhan Before 1998	.00	.12	−.12
		(.03)	(.04)
Took Part in Panchayat Activities	.28	.78	−.50
Prior to Being Elected	(.06)	(.04)	(.07)
Knew How GP Functioned	.00	.35	−.35
		(.05)	(.07)
Did Not Receive any Formal Training	.06	.00	.06
	(.03)		(.02)
Spouse ever Elected to the Panchayat	.17	.02	.15
	(.05)	(.01)	(.04)
Spouse Helps	.43	.13	.30
	(.07)	(.03)	(.07)
Will Not Run Again	.33	.21	.13
	(.06)	(.04)	(.07)
C. *Pradhan's Political Party*			
Left Front	.69	.69	−.01
	(.06)	(.04)	(.08)
Right (Trinamul or BJP)	.19	.18	.01
	(.05)	(.04)	(.06)
Observations	54	107	

Note: 1. Standard errors, corrected for clustering at the GP level using the Moulton (1986) formula, are in parentheses.

1 Event Studies

- Used mostly in finance to look at impact of events on asset prices

- Can be used in other contexts. Need:
  - High frequency data on dependent variable
  - Rare large changes in "independent variable" ("event")

- Examples:
  - Impact of Election on Stock Prices
  - Impact of Leader Death on Stock Prices
  - Impact of Unions on Health Care Quality (mortality, wound infections, urinary tract infections, etc.)

- Two Time Periods:
  - Estimation Window (Normal Times): time τ_0 to τ_1
  - Event Window (Special Times): time $\tau_1 + 1$ to τ_2
    * 0 is date of event
    * $0 \in [\tau_1 + 1, \tau_2]$
- Estimate a model of outcome in Estimation Window:
  - $Y_{it} = f(X_{it}\beta) + \epsilon_{it}$
  - Assume $\epsilon_{it}\tilde{}N\left(0, \sigma^2_{\epsilon_i}\right)$
  - Estimate $\widehat{\beta}$
- Compute Abnormal Returns (AR) in Event Window

- $AR_{it} = Y_{it} - f\left(X_{it}\widehat{\beta}\right)$

- Model for Abnormal Returns
  - Constant Mean Model: $f(X_{it}\beta) = \mu_i$
  - Market Model: $f(X_{it}\beta) = \alpha_i + \beta_i R_{mt}$ (where R_{mt} is the value of a market index)
  - Factor Models: $f(X_{it}\beta) = \alpha_i + Z_{it}\beta_i$
  - CAPM, APT

- Cumulative Abnormal Returns (CARS)
  - $CAR_{it} = \sum_{t=\tau_1+1}^{\tau_2} AR_{it}$ (sum of abnormal returns over event window)

- Estimating CARS (example - market model):

– Regression Method

  * $Y_{it} = \alpha_i + \beta_i R_{mt} + \gamma_i D_t + \epsilon_{it}$

  * where $D_t = 1$ in the event window and 0 in the estimation window

– Summing Up Method

  * Estimate $\hat{\beta}_i$ out of sample and then $CAR_{i\tau} = \sum_{t=\tau_1+1}^{\tau_2} Y_{it} - f\left(X_{it}\widehat{\beta}\right)$

  * Example with market model: $AR_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt}$

$$
\begin{aligned}
& V\left(AR_{i\tau} | R_{mt}\right) \\
= \; & \sigma^2_{\epsilon_i} + \frac{1}{\tau_2 - \tau_1 - 1}\left[1 + \frac{\left(R_{m\tau} - \hat{\mu}_m\right)^2}{\hat{\sigma}^2_m}\right]
\end{aligned}
$$

- In-Sample Versus Out of Sample Estimation

– In-Sample: Effect of event can impact estimation. For instance, it can impact estimation of variances $\sigma^2_{\epsilon_i}$.

– Out-of-Sample: Implicitly assume that structure of variances/covariances (higher moments) are the same in and out of sample. Therefore, the null hypothesis is really a more general hypothesis that the distribution is the same.

– Out-of-Sample: Adding covariates always helps with in sample fit. By definition (since one can always set $\gamma = 0$ and achieve the same residual variance:

$$
\begin{aligned}
& \min_{\beta} \sum_{t=\tau_1+1}^{\tau_2} (R_{it} - X_{it}\beta)^2 \\
\geq \; & \min_{\beta,\gamma} \sum_{t=\tau_1+1}^{\tau_2} (R_{it} - X_{it}\beta - Z_{it}\gamma)^2
\end{aligned}
$$

However, adding Z for out of sample fit can increase prediction error and thus abnormal return variance. From the above expression of variance, we can see that this is less true when there is a large sample size.

* Consequence: Event studies (which tend to rely on summing up method) tend to use few covariates

* Interpretation: Essentially the prediction technique does not take into account the variance of the estimate. If I take any variable, it is likely to be correlated with my independent variable in small samples. If the correlation is purely random, the coefficients may be large but should have large standard errors. Essentially good prediction techniques should account for this and weight by the variance of the estimate.

3 Estimating the Value of Political Connections

- One of the first papers in "Forensic Economics"

- Looks at stock returns for companies connected to Suharto in Indonesia on when Suharto had negative health shocks

- Uses Event Study Methodology

- Measure of connectedness from Suharto Dependency Index (1995) by the Castle Group (leading consulting firm in Indonesia) - ranks companies 1-5 where 5 is most dependent on Suharto.

  - Total of 25 industrial groups

  - Sample of 79 firms

- Obtained dates of health events from doing a keyword search on Lexis Nexis (Suharto, Health, Indonesia) and (Stock or Financial)

- Obtained boundary of dates by???

- Estimates separately for each event:

$$R_{ie} = \alpha + \rho POL_i + \epsilon_{ie} \tag{4}$$

$$\begin{aligned} R_{ie} &= \alpha + \rho_1 POL_i + \rho_2 NR_e(JCI) \\ &\quad + \rho_3 \left[NR_e(JCI) * POL_i \right] + \epsilon_{ie} \end{aligned} \tag{5}$$

where R_{ie} is the return on security i during episode e, POL_i is the political connectedness of firm i and $NR_e(JCI)$ is a broad South East Asia market index during episode e

- Also does robustness

– Allows for non-parametric functional form in POL variable

– Used beginning of SE Asian Crisis as robustness to whether connected firms are just more vulnerable to bad news (problem: SE crisis also lead to downfall of Suharto)

– Restricted sample only to firms with active trading during events (i.e. firms with thin trading might have "effect" but just not get traded)

– Looked at runup before events (i.e. insider trading)

TABLE 1—SUMMARY STATISTICS BY DEGREE OF POLITICAL DEPENDENCE AS MEASURED BY THE SUHARTO DEPENDENCY INDEX

POL	1	2	3	4	5	All firms	Observations
Observations	5	34	10	16	14	79	
Assets	2,145.76 (2,843.63)	2,228.57 (3,989.85)	2,206.20 (3,676.99)	1,634.08 (2,561.07)	1,765.51 (2,230.52)	2,033.19 (3,321.59)	76
Debt	707.18 (702.84)	791.32 (1,478.83)	813.25 (976.28)	397.83 (461.06)	712.57 (1,070.83)	717.37 (1,186.85)	70
Return on assets (net income)/(total assets)	0.038 (0.031)	0.058 (0.058)	0.043 (0.023)	0.037 (0.032)	0.050 (0.029)	0.050 (0.044)	76
Tax rate (taxes paid)/(pretax income)	0.23 (0.05)	0.24 (0.12)	0.16 (0.14)	0.22 (0.16)	0.15 (0.12)	0.21 (0.13)	74

Sources: All data are from the *Financial Times' Extel Database* (1997); Assets and Debt are expressed in millions of 1995 rupiah.

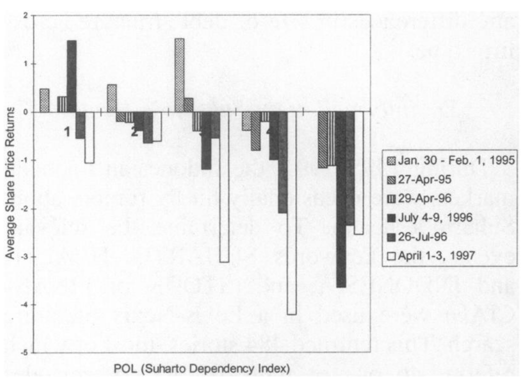


FIGURE 1. EFFECT OF POLITICAL DEPENDENCE ON SHARE PRICE RETURNS

ascertain the date when rumors first hit the Jakarta Exchange—there was generally a specific triggering *event*, which I take as the start of the episode. I assumed that each episode came to an end when it was (1) explicitly put to rest by the revelation of new information or (2) it was reported that analysts had factored the new information about Suharto's health into their pricing of securities.

II. Results

Figure 1 shows the share price returns for the six episodes, with the Suharto Dependency Index on the horizontal axis. The graph strongly suggests that politically dependent firms, on average, lost more value during these episodes than did less-dependent firms.

To get a sense of the magnitude of the effect of political dependence during each episode, I ran a set of regressions using the following specification:

$$R_{ie} = \alpha + \rho \cdot POL_i + \varepsilon_{ie} \tag{1}$$

where R_{ie} is the return on the price of security i during episode e, POL_i is the firm's Suharto Dependency Number, and ε_{ie} is the error term.[5] The results of this set of regressions are listed in Table 2; consistent with the raw pattern illustrated in Figure 1, ρ is negative in every instance.

Now, in each episode, investors were reacting to a different piece of news, so we expect the coefficient on POL_i to differ across events. More precisely, a more severe threat to Suharto's health should intensify the effect of political dependence, hence the magnitude of ρ should be increasing with event severity. As a measure of the market's concerns regarding the threat to Suharto's health in each episode, I use

[5] All regressions reported in this paper use standard errors that correct for heteroskedasticity. I also ran regressions using an error structure that only allowed for the correlation of ε_{ei}s for each company, i.e., $\mathrm{Cov}(\varepsilon_{ei}, \varepsilon_{fj}) \neq 0$ if and only if $i = j$. The regressions were also run using an error structure that allowed for the correlation of ε_{ei}s within each group. These various approaches yielded very similar sets of standard errors.

TABLE 2—EFFECT OF POLITICAL CONNECTIONS ON CHANGES IN SHARE PRICE, SEPARATE ESTIMATION FOR EACH EVENT

	Jan. 30–Feb. 1, 1995	April 27, 1995	April 29, 1996	July 4–9, 1996	July 26, 1996	April 1–3, 1997
POL	−0.58* (0.34)	−0.31 (0.18)	−0.24* (0.15)	−0.95*** (0.27)	−0.57*** (0.22)	−0.90** (0.35)
Constant	1.29 (0.79)	0.21 (0.32)	0.12 (0.46)	0.83 (0.64)	−0.07 (0.41)	0.77 (0.97)
R^2	0.037	0.043	0.025	0.147	0.078	0.075
Observations	70	70	78	79	79	79

Note: Robust standard errors are in parentheses.
* Significantly different from 0 at the 10-percent level.
** Significantly different from 0 at the 5-percent level.
*** Significantly different from 0 at the 1-percent level.

the return on the Jakarta Stock Exchange Composite Index net of broader Southeast Asian effects[6] [referred to using $NR_e(JCI)$]. The preceding observations suggest that the coefficient on *POL* should be more negative if the threat to Suharto's health, as proxied by $NR_e(JCI)$, is greater.[7] This turns out to be the case: the correlation between ρ and $NR_e(JCI)$ is 0.98. This implies a specification where observations from all events are pooled together, with an interaction term, $NR_e(JCI) * POL_i$, added to allow the effect of political dependence to vary across events, depending on the event's severity. Thus, I use the following full-sample specification:

$$(2)\quad R(P_{ie}) = \alpha + \rho_1 \cdot POL_i + \rho_2 \cdot NR_e(JCI) + \rho_3 \cdot [NR_e(JCI) \cdot POL_i] + \varepsilon_{ie}.$$

The results of this regression are listed in Table 3.[8]

TABLE 3—EFFECT OF POLITICAL CONNECTIONS ON CHANGES IN SHARE PRICE

	(1)	(2)
POL	−0.60** (0.11)	−0.19 (0.15)
NR(JCI)	0.25 (0.14)	−0.32 (0.28)
NR(JCI) · POL		0.28* (0.11)
Constant	0.88 (0.27)	0.06 (0.35)
R^2	0.066	0.078
Number of observations	455	455

Note: Robust standard errors are in parentheses.
* Significantly different from 0 at the 5-percent level.
** Significantly different from 0 at the 1-percent level.

If the severity of a rumor affects politically dependent firms more than less-dependent firms, then the coefficient on the interaction term $NR_e(JCI) \cdot POL_i$ should be positive. The estimated coefficient, ρ_3, is statistically significant at 5 percent and is equal to 0.28. Thus, if the overall market declined by 1 percent in reaction to news about Suharto's health, we might expect a firm with $POL = x$ to drop 0.28 percent more than a firm with $POL = x - 1$.

[6] To net out broader Southeast Asia effects, I ran the following "market model" for daily returns during 1994:

$$R_t(JCI) = \alpha + \sum_{m \in M} \beta_m \cdot R_t(m) + \varepsilon_t$$

where $R_t(JCI)$ is the return on the Jakarta Composite on day t, $R_t(m)$ is the return on market index m, and M is the set of ASEAN market indices (including Tokyo's Nikkei 225, Hong Kong's Hang Seng, Singapore's Straits Times, Bangkok's SET, Taiwan's Weighted, Philippines' Composite, Kuala Lumpur's Composite, and Seoul's Composite). This produced a set of coefficients reflecting the degree of correlation between the JCI and other market indices. For each episode e, the *net* return for the JCI is then given by

$$NR_e(JCI) = R_e(JCI) - [\hat{\alpha} + \sum_{m \in M} \hat{\beta}_m \cdot R_e(m)].$$

[7] It may seem somewhat circular to use $NR_e(JCI)$ as a measure of the severity of the threat to Suharto's health when many of the firms in my sample are constituents of the JCI. Note, however, that $NR_e(JCI)$ is a difference, of which the coefficient on *POL* is a difference in differences. As Section III, subsection B, illustrates, these two variables need not be correlated.

[8] Regressions were also run using log(ASSETS), log(DEBT), and industry dummies as controls. These additions did not alter the size of significance of the interaction term.

1 Gauss Markov Assumptions

- OLS is minimum variance unbiased (MVUE) if
  - Linear Model: $Y_i = X_i \beta + \epsilon_i$
  - $E\left(\epsilon_i | X_i\right) = 0$
  - $V\left(\epsilon_i | X_i\right) = \sigma^2 < \infty$
  - $cov\left(\epsilon_i, \epsilon_j\right) = 0$
  - Normally distributed errors.

- What happens if we relax homoskedasticity? Uncorrelated errors?
  - Bias of $\hat{\beta}$? No!
  - Bias of $SE\left(\hat{\beta}\right)$?
    * Yes, distorted test size: OLS formula for standard errors not valid: $\sigma^2 \left(X'X\right)^{-1}$

    * Up or down? Could be either (In general, positive correlation $\implies$ OLS standard errors are too low, negative correlation $\implies$ OLS standard errors are too high).

  - OLS not MVUE anymore

- This lecture will be about what to do when the homoskedasticity and uncorrelated errors assumptions are relaxed

2 Non-Spherical Disturbances: Examples

2.1 Classical OLS

$$\begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

2.2 Heteroskedasticity

$$\begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{pmatrix}$$

2.3 General

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \sigma_{45} \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{45} & \sigma_5^2 \end{pmatrix}$$

2.4 General Clustered (with G clusters)

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & . & . & 0 & 0 & 0 \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & . & . & 0 & 0 & 0 \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & . & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ 0 & 0 & 0 & . & . & \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ 0 & 0 & 0 & . & . & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}$$

2.5 Random Effects Model

- Each cluster is structured as

$$\begin{pmatrix} \sigma^2 + \sigma_G^2 & \sigma_G^2 & \sigma_G^2 \\ \sigma_G^2 & \sigma^2 + \sigma_G^2 & \sigma_G^2 \\ \sigma_G^2 & \sigma_G^2 & \sigma^2 + \sigma_G^2 \end{pmatrix}$$

2.6 Clustered AR(1) Model

$$\begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & . & . & 0 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & . & . & 0 & 0 & 0 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & . & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ 0 & 0 & 0 & . & . & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & . & . & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

3 Bias in Standard Errors with Non-Spherical Disturbances

- Model Outline: Assume
  - $Y = X\beta + \epsilon$
  - $V(X) = \sigma_X^2$
  - $V(\epsilon) = \sigma_\epsilon^2$
  - $Cov\left(X_{itg}, X_{isg}\right) = \rho_x$
  - $Cov\left(X_{itg}, X_{itg'}\right) = 0$
  - $Cov\left(\epsilon_{itg}, \epsilon_{isg}\right) = \rho_\epsilon$
  - $Cov\left(\epsilon_{itg}, \epsilon_{itg'}\right) = 0$
- OLS

- $\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$
- $SE\left(\hat{\beta}_{OLS}\right) = (X'X)^{-1} (X'\Omega X) (X'X)^{-1}$

- Note that
  - $X\prime X = \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^2$
  - $X'\epsilon = \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}\epsilon_{it}$
  - Since X is one dimensional vector, we get

$$\begin{aligned} SE\left(\hat{\beta}_{OLS}\right) &= \left(X'X\right)^{-1} \left(X'\Omega X\right) \left(X'X\right)^{-1} \\ &= \left(X'X\right)^{-2} X'\Omega X \end{aligned}$$

*

$$\Longrightarrow \quad p\lim SE\left(\hat{\beta}_{OLS}\right) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^2\right)^{-2} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}\epsilon_{it}\right)^2$$

$$= \frac{NT\sigma_X^2\sigma_\epsilon^2 + NT\,(T-1)\,\rho_X\rho_\epsilon}{\left(NT\sigma_X^2\right)^2}$$

$$= \frac{\sigma_\epsilon^2 + (T-1)\,\frac{\rho_X\rho_\epsilon}{\sigma_X^2}}{NT\sigma_X^2}$$

– Implications:

  * $\rho_x > 0, \rho_\epsilon > 0 \implies$ OLS standard errors downward biased: interpretation - some of the lack of variation is not independent

  * $\rho_x > 0, \rho_\epsilon < 0 \implies$ OLS standard errors upward biased: interpretation - some of the variation is not independent

4 Three Types of Fixes

- Keep $\hat{\beta}$ estimate and adjust standard errors.
  - Eicker-White heteroskedasticity robust standard errors
  - Cluster-Robust standard errors (called "clustering the standard errors")
  - Use complete variance-covariance matrix for inference
- Alter the estimator of $\hat{\beta}$ in addition to using non-OLS standard errors
  - GLS - Generalized Least Squares
  - FGLS - Feasible Generalized Least Squares
  - MLE - Maximum Likelihood
- Collapse data

5 General Tradeoff

- By imposing structure you get greater efficiency
  - Less parameters to estimate
  - More observations per parameter
- But you could be wrong about the structure in which case you could have the wrong standard errrors

6 Eicker-White Heteroskedasticity Robust Standard Errors

- Heteroskedasticity robust standard errors keeps the OLS estimator but changes the standard errors by using the formula

$$V\left(\hat{\beta}_{OLS}\right) = \left(X'X\right)^{-1} X'\hat{\Omega}X \left(X'X\right)^{-1}$$

where $\hat{\Omega} =$

$$\begin{pmatrix} \hat{\epsilon}_1^2 & 0 & 0 & 0 & 0 \\ 0 & \hat{\epsilon}_2^2 & 0 & 0 & 0 \\ 0 & 0 & \hat{\epsilon}_3^2 & 0 & 0 \\ 0 & 0 & 0 & \hat{\epsilon}_4^2 & 0 \\ 0 & 0 & 0 & 0 & \hat{\epsilon}_5^2 \end{pmatrix}$$

- In other words:

$$V\left(\hat{\beta}_{OLS}\right) = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{N} \hat{\epsilon}_i^2 x_i x_i'\right) \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$$

- Note that the sample size for estimating σ_i^2 is one so that we do not have a consistent estimate of σ_i^2.

- Tradeoff with GLS

  - Negative: Less efficient if truly heteroskedastic

  - Positive: Doesn't require knowledge of the variance-covariance matrix

7 Clustered Standard Errors

- When error terms are correlated within groups but not across groups and when the division of observations into groups is known, standard errors can be "clustered" or adjusted for within-group correlation.

- Clustered standard errors allow for arbitrary patterns of correlation within clusters (groups). Many clusters are needed to invoke assymptotic approximations (Donald and Lang, 2007).

7.1 Single Dimensional Clustering

Cluster-robust standard errors formula:

$$\left(X'X\right)^{-1} X'\hat{\Omega}X \left(X'X\right)^{-1}$$

where $\hat{\Omega} =$

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & . & . & 0 & 0 & 0 \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & . & . & 0 & 0 & 0 \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & . & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ 0 & 0 & 0 & . & . & \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ 0 & 0 & 0 & . & . & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}$$

In other words:

$$V\left(\hat{\beta}_{OLS}\right) = \left(\sum_{c=1}^{C} X_c' X_c\right)^{-1} \sum_{c=1}^{C} X_c' \hat{\epsilon}_c \hat{\epsilon}_c' X_c \left(\sum_{c=1}^{C} X_c' X_c\right)^{-1}$$

7.2 Multi-Dimensional Clustering

- Suppose correlation exists in multiple dimensions within two dimensions of groups over time (i.e. within workers over time and across workers within a certain block of time)

- Two options:

  - Choose one dimension relevant to the parameter of interest and cluster only on one dimension

  - Cluster on two dimensions

- Assumptions

  - $Y_{ijt} = X_{ijt}\beta + \epsilon_{ijt}$

  - $cov\left(\epsilon_{ijt}, \epsilon_{kjt}\right) \neq 0$

  - $cov\left(\epsilon_{ijt}, \epsilon_{imt}\right) \neq 0$

  - $cov\left(\epsilon_{ijs}, \epsilon_{ijt}\right) = 0$

- So:

$$\begin{aligned} V\left(\hat{\beta}_{2D}\right) &= \left(X'X\right)^{-1}\hat{Q}\left(X'X\right)^{-1} \\ \text{where } \hat{Q} &= X'\left(\hat{\Omega}S^{IJ}\right)X \end{aligned}$$

- $S^{IJ} = S^I + S^J - S^{I \cap J}$ where S^K is the cluster matrix for dimension K

$$\begin{aligned}\hat{Q} &= X'\left(\hat{\epsilon}\hat{\epsilon}' S^{IJ}\right) X = \\ & X'\left(\hat{\epsilon}\hat{\epsilon}' S^{I}\right) X + X'\left(\hat{\epsilon}\hat{\epsilon}' S^{J}\right) X - \\ & X'\left(\hat{\epsilon}\hat{\epsilon}' S^{I \cap J}\right) X\end{aligned}$$

  - A cluster matrix is a matrix of zeros and ones where a zero is entered if the entry in the variance-covariance matrix is assumed to be zero and a one is entered if the entry in the variance-covariance matrix is estimated. Example: Let S^I be given by (consecutive groupings):

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

– and S^J be given by (odd and even groupings):

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

– Then, the intersection matrix enters a one if entries from both cluster matrices (S^I and S^J) are one and zero otherwise:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- So $V\left(\hat{\beta}_{2D}\right) =$

$$\left(X'X\right)^{-1} X'\left(\hat{\epsilon}\hat{\epsilon}' S^I\right) X \left(X'X\right)^{-1} +$$
$$\left(X'X\right)^{-1} X'\left(\hat{\epsilon}\hat{\epsilon}' S^J\right) X \left(X'X\right)^{-1} -$$
$$\left(X'X\right)^{-1} X'\left(\hat{\epsilon}\hat{\epsilon}' S^{I\cap J}\right) X \left(X'X\right)^{-1}$$

- Thus, estimate 3 separate OLS regressions: one clustered by S^I, the next by S^J, and the third by $S^{I\cap J}$ and then compute the abve formula.

8 Weighted Least Squares

- We now introduce estimators where we alter the estimation of β in addition to the standard errors. Why would we do this? Efficiency!

8.1 GLS

- Estimation
  - Variance-covariance matrix known: Ω

– Regress $\Omega^{-\frac{1}{2}}Y = \Omega^{-\frac{1}{2}}X\beta + \Omega^{-\frac{1}{2}}\mu$

– $\hat{\beta} = \left(X'\Omega X\right)^{-1} X'\Omega Y$

– Downweights high variance observations, upweights low variance observations

– Takes into account cross-observation correlation patterns

- Positive
  – Can handle arbitrary correlation structures
  – Efficient if you know the correlation structure

- Negative
  – Relies on knowing the variance-covariance matrix Ω

- Weights efficiently so doesn't estimate average treatment effect in the presence of treatment effect heterogeneity

8.2 FGLS

- Estimation
  - Stage 1: Run OLS - $Y = X\beta + \mu$
  - Stage 2: extract variance-covariance matrix from stage 1 - $\hat{\Omega}$ and run GLS with estimated matrix: Regress $\hat{\Omega}^{-\frac{1}{2}}Y = \hat{\Omega}^{-\frac{1}{2}}X\beta + \hat{\Omega}^{-\frac{1}{2}}\mu$
  - $\hat{\beta} = \left(X'\hat{\Omega}X\right)^{-1} X'\hat{\Omega}Y$
- Positive
  - Can handle arbitrary correlation structures

– Doesn't rely on knowing the variance-covariance matrix Ω

- Negative

  – Biased in small samples: $E\left(X'\hat{\Omega}X\right)^{-1}X'\hat{\Omega}Y \neq \beta$

  – Variance-covariance matrix noisy. Note that $\hat{\beta}_{FGLS}$ is consistent for β but $\hat{\Omega}$ is not consistent for Ω. To estimate, $\hat{\Omega}$, we need to estimate $\frac{N(N+1)}{2}$ entries of the variance-covariance matrix for a sample size of N

  – Weights efficiently so doesn't estimate average treatment effect in the presence of treatment effect heterogeneity

9 Maximum Likelihood

- Can structurally model error terms - easy to allow for non-spherical disturbances

  - Note: not all distributions have an independent variance parameter - some like poisson, negative binomial, exponential have only one parameter. Others like the normal, lognormal have independent mean and variances.

- Benefits of MLE

  - Can have better small sample properties if you know the error term

  - Easier to model error structure

  - Reachers Cramer-Rao lower bound - efficient!

- Costs

– You need to know the distribution

– Not consistent if the distribution is wrong

– Can be biased in small samples even if the distribution is correct

– Doesn't generally have closed form computational formulas - have to solve simultaneously for set of first order conditions. Additional problems of knowing whether a solution to the set of first order conditions is a local/global maximum/minimum.

10 Structured FGLS:

10.1 Example - Cochrane-Orcutt

- Assume $Y_{it} = X_{it}\beta + \epsilon_{it}$ where $\epsilon_{it} = \rho\epsilon_{it-1} + \mu_{it}$

- Then follow these steps:

    1. Estimate $Y_{it} = X_{it}\beta + \epsilon_{it}$

    2. Regress $\epsilon_{it} = \rho\epsilon_{it-1} + \mu_{it}$ and obtain ρ

    3. Then transform data to

    $$Y_{it} - \rho Y_{it-1} = (X_{it} - \rho X_{it-1})\,\beta + \epsilon_{it}$$

    4. $\hat{\beta}$ is now correct and so are the OLS standard errors

10.2 Example: Newey-West

- Variance covariance matrix with each cluster assumed to equal:

$$\begin{pmatrix} \sigma^2 & \left[1-\frac{1}{M}\right]\sigma^2 & . & \left[1-\frac{K}{M}\right]\sigma^2 \\ \left[1-\frac{1}{M}\right]\sigma^2 & \sigma^2 & . & \left[1-\frac{K-1}{M}\right]\sigma^2 \\ \left[1-\frac{2}{M}\right]\sigma^2 & \left[1-\frac{1}{M}\right]\sigma^2 & . & \left[1-\frac{K-2}{M}\right]\sigma^2 \\ . & . & . & . \\ \left[1-\frac{K}{M}\right]\sigma^2 & \left[1-\frac{K-1}{M}\right]\sigma^2 & . & \sigma^2 \end{pmatrix}$$

- The above formulation, called the Newey-West estimator, allows for linear fall off in correlation of error terms within clusters

- Can be estimated using GLS or MLE

11 Collapsing

- Suppose that X variables are the same within cluster so that

$$Y_{ig} = \alpha + \beta X_g + C_g + \epsilon_{ig}$$

- Then there is no loss in collapsing the data because there is no within cluster variation used to identify β

- Otherwise you trade off:

  – Not using variation from a correlation structure you do not know

  – Throwing away useful correlation within clusters from covariates X