# Causal Inference with Time-Series Cross-Sectional Data: with Applications to Positive Political Economy

by

Yiqing Xu

M.A. in Economics

Peking University, 2010



SUBMITTED TO THE DEPARTMENT OF POLITICAL SCIENCE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUNE 2016



Signature of Author ......... Signature redacted ................

Department of Political Science

April 14, 2016

Certified by .......... Signature redacted ......................

Teppei Yamamoto

Associate Professor of Political Science

Thesis Supervisor

Accepted by ....... Signature redacted .............................

Ben Ross Schneider

Ford International Professor of Political Science

Chair, Graduate Program Committee

# Causal Inference with Time-Series Cross-Sectional Data: with Applications to Positive Political Economy

by

Yiqing Xu

Submitted to the Department of Political Science on April 14, 2016
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Political Science

**ABSTRACT**

Time-series cross-sectional (TSCS) data are widely used in today's social sciences. Researchers often rely on two-way fixed effect models to estimate causal quantities of interest with TSCS data. However, they face the challenge that such models are not applicable when the so called "parallel trends" assumption fails, that is, the average treated counterfactual and average control outcome do not follow parallel paths.

The first chapter of this dissertation introduces the *generalized synthetic control* method that addresses this challenge. It imputes counterfactuals for each treated unit using control group information based on a linear interactive fixed effect model that incorporates unit-specific intercepts interacted with time-varying coefficients. It not only relaxes the often-violated "parallel trends" assumption, but also unifies the synthetic control method with linear fixed effect models under a simple framework.

The second chapter examines the effect of Election Day Registration (EDR) laws on voter turnout in the United States. Conventional difference-in-differences approach suggests that EDR laws had almost no impact on voter turnout. Using the generalized synthetic control method, I show that EDR laws increased turnout in early adopting states but not in states that introduced them more recently.

The third chapter investigates the role of informal institutions on the quality of governance in the context of rural China. Using TSCS analysis and a regression discontinuity design, I show that village leaders from large lineage groups are associated with considerably more local public investment. This association is stronger when the groups appeared to be more cohesive.

Thesis Supervisor: Teppei Yamamoto
Title: Associate Professor of Political Science

# Acknowledgment

It is hard to believe that the six-year long journey at M.I.T. is coming to an end. I still remember the excitement when I first came to Cambridge and visited M.I.T. As I was walking on the bridge connecting E52 and E62, I felt that I was so lucky to able to spend the next five, six years in this place.

Navigating through this journey has not been easy. There is simply too much to learn and there are too many things to choose from. This thesis is not only a collection of work that I have finished at M.I.T., but also represents things I have learned over the years. It would not be possible without the support of my teachers and friends, whom I would like to acknowledge here.

First and foremost, I would like to thank Professor Teppei Yamamoto, Chair of my dissertation committee. It is my honor to be his first official PhD student. Teppei is a great mentor, always caring and thoughtful. He guided me through the last three years of PhD study step by step. He spent hours with me standing in front of a whiteboard, helping me with math problems. He read my papers line by line and gave me suggestions so meticulously that I felt guilty about my sloppiness. He revised the abstract of a paper of mine over and over again and in the end I was amazed how much improvement we had made. I feel much indebted to him. He taught me not only knowledge, but also skills and know-hows required by an independent researcher.

I would like to thank Professor Jens Hainmueller, who has been extraordinarily supportive all the way along. He was very reason I began to think about the main theme of this dissertation and was always there to offer help when I encountered difficulties. Jens genuinely cares about his students and I am very lucky to be one of them. I visited Stanford last year because of him and benefited tremendously from this experience. After I received an invitation for a job interview, Jens checked with me almost everyday, giving me advice and helping me revise every single page of my slides. Working with him was also a lot of fun. I wish one day I could write and communicate as well as he does.

I also wish to thank Professors Lily Tsai and Jim Snyder, who have taught and helped me immensely. Lily was my academic advisor in the first two years. I really appreciate the freedom she gave me in terms of choosing classes and finding a path of my own. Mainly thanks to her, I recognized my huge knowledge gap in the political science literature (and especially in comparative politics). Lily is always there, helping me close that gap. I am still not there yet, but I will always keep in mind her suggestions. Jim is a scholar I admire even before I came to M.I.T. He convinced me to come to Cambridge and M.I.T. six years ago and I never regretted this choice. The PE breakfast he used to run at M.I.T. was one of the most effective venues in which we learned how to find a good research question, how to design and execute a research project, how to present, and how to criticize each other's work. It was more challenging than most conferences and seminars we later attend.

I would like to thank my friends, teachers, and colleagues, from whom I learned a great deal. I would especially like to thank my coauthors Jennifer Pan and Jidong Chen, who are now professors at leading universities in the U.S. and China, and Professors Devin Caughey and Chris Warshaw at M.I.T., who helped me expand my scope of research. I wish to thank Professors Danny Hidalgo and In Song Kim, whom I very much enjoy working with as a teaching assistant. I have always benefited from conversations with Professors Yasheng Huang at M.I.T. Sloan and Rod MacFarquhar and Liz Perry at Harvard Government.

Finally, I want to thank my parents for supporting me throughout this journey. You give me everything and I am sorry that I have not been quite around during the past few years.

I will miss this town.

*This thesis is supported by a doctoral fellowship provided by the Chiang Ching-kuo Foundation (North America). I thank the Foundation for its professionalism and generosity.*

# Table of Contents


Abstract ........................................................................................................ 3

Acknowledgment ............................................................................................ 5

Introduction .................................................................................................... 9

Chapter 1 .......................................................................................................13

Chapter 2 .......................................................................................................43

Chapter 3 .......................................................................................................75

# Introduction

One of the main goals of social science inquiries is to establish causal relationships between real world phenomena. The primary method of establishing causalities is randomized controlled experiments. However, under many circumstances, researchers are unable to conduct experiments because of resource constraints and/or ethical concerns; nor do they always have the opportunity to exploit natural experiments. Previous researches have shown that employing designs with observational, longitudinal data, data of both space and time dimensions, is a promising way to establish causality. This thesis attempts to improve methods of causal inference with time-series cross-sectional (TSCS) data—longitudinal data with many time periods—and apply them to answer important real-world questions.

The core and methodological part of this thesis is what I call the generalized synthetic control method. It unifies fixed effects models, including difference-in-differences (DID), and the synthetic control mothed under a single framework. Specifically, it improves causal inference with TSCS data when treatment units are not randomly selected. Such cases are ample in political science. For example, researchers might be interested in the effect of a reform that took place in several US states, but those states could be fundamentally different from the rest of the country. Conventional two-way fixed effect models may not be useful when the average treated counterfactual and average control outcome do not follow parallel paths (in which case the so called "parallel trends" assumption fails).

The generalized synthetic control method addresses this challenge. The basic idea is to take into account unobserved time-varying confounders by decomposing the error structure into lower-dimensional factors and conditioning on these factors. It imputes counterfactuals for each treated unit using control group information based on a linear interactive fixed effect model that incorporates unit-specific intercepts interacted with time-varying coefficients. This method is in the spirit of the original synthetic control method in the sense that, like synthetic control, it uses pre-treatment periods to learn the relationships between treated and control units, based on which it predicts counterfactuals for each treated unit.

This method is widely applicable in political science. For instance, researchers can use this method to estimate the effect a country's joining an international organization on its probability of having conflicts with other countries, or to examine the effect of foreign aid on economic growth. In both cases, we cannot readily assume the parallel trends assumption to be valid. This method has several attractive features. First, because it allows the treatment to be correlated with unobserved unit and time heterogeneities, it is more robust and often more efficient than conventional fixed-effect models. Second, it generalizes the synthetic control method to the case of multiple treated units and variable treatment timing. With this method, users no longer need to find matches for each treated unit since the algorithm produces treated counterfactuals in a single run. Moreover, it addresses the inferential problem of the original synthetic control method and gives more interpretable uncertainty estimates. Finally, with a built-in cross-validation procedure, it avoids specification searches and thus is easy to implement.

The second chapter of this thesis apply the generalized synthetic control method to an empirical application in American politics. Previous researches have not reached a consensus whether Election Day Registration (EDR), a reform meant to reduce the cost of voting, contributed to an increase in voter turnout. The difficulty of causal identification lies in the fact that states that have adopted EDR laws are intrinsically different from those that have not. The two groups of states do not share parallel paths in the pre-EDR law era, suggesting that a DID approach is not a valid identification strategy. Using the new method, I find that EDR laws increased voter turnout in early adopting states, but not in states that introduced EDR as a strategy to opt out the 1993 National

Voter Registration Act or enacted EDR laws in recent years. These results are broadly consistent with evidence provided by a large literature based on individual-level cross-sectional data. They are also more credible than results from conventional fixed effects models when the "parallel trends" assumption appears to fail.

In the third chapter, I apply TSCS analysis to answer an important empirical question in comparative politics: Do informal institutions matter for local governance in environments of weak democratic or bureaucratic institutions? This question is difficult to answer because of challenges in defining and measuring informal institutions and identifying their causal effects. In the context of rural China, I investigate the effect of lineage groups on local public goods expenditure. Using a TSCS dataset of 220 Chinese villages from 1986 to 2005, I find that village leaders from the two largest family clans in a village increased local public investment considerably. This association is stronger when the clans appeared to be more cohesive. I also find that clans helped local leaders overcome the collective action problem of financing public goods, but there is little evidence suggesting that they held local leaders accountable.

# Chapter 1

The first chapter of this thesis proposes a new method for causal inference with time-series cross-sectional (TSCS) data, which I call the generalized synthetic control (GSC) method. First, I discuss the literature and provide motivations for developing a new method for causal inference with TSCS data. Then I set up the model and define the main quantity of interest, after which I introduce the GSC estimator and its implementation procedure. The following section discusses the inferential method for the GSC estimator. The last section provides a short summary. Simulation results, additional robustness checks, as well as an empirical example, are provided in the next chapter.

## Motivation

Difference-in-differences (DID) is one of the most commonly used empirical designs in today's social sciences. The identifying assumptions for DID include the "parallel trends" assumption, which states that in the absence of the treatment the average outcomes of treated and control units would have followed parallel paths. This assumption is not directly testable, but researchers have more confidence in its validity when they find that the average outcomes of the treated and control units follow parallel paths in pre-treatment periods. In many cases, however, parallel pre-treatment trends are not supported by data, a clear sign that the "parallel trends" assumption is likely to fail in the post-treatment period as well. This paper attempts to deal with this problem systematically. It proposes a method that estimates the average treatment effect on the treated using time-series cross-sectional (TSCS) data when the "parallel trends" assumption is not likely to hold.

The presence of unobserved time-varying confounders causes the failure of this assumption. There are broadly two approaches in the literature to deal with this problem. The first one is to condition on pre-treatment observables using matching methods, which may help balance the influence of potential time-varying confounders between treatment and control groups. For example, Abadie (2005) proposes matching before DID estimations. Although this method is easy to implement, it does not guarantee parallel pre-treatment trends. The synthetic control method proposed by Abadie, Diamond and Hainmueller (2010, 2015) goes one step further. It matches both pre-treatment covariates and outcomes between a treated unit and a set of control units and uses pre-treatment periods as criteria for good matches.[1] Specifically, it constructs a "synthetic control unit" as the counterfactual for the treated unit by reweighting the control units. It provides explicit weights for the control units, thus making the comparison between the treated and synthetic control units transparent. However, it only applies to the case of one treated unit and the uncertainty estimates it offers are not easily interpretable.[2]

The second approach is to model the unobserved time-varying heterogeneities explicitly. A widely used strategy is to add in unit-specific linear or quadratic time trends to conventional two-way fixed effects models. By doing so, researchers essentially rely upon a set of alternative identification assumptions that treatment assignment is ignorable conditional on both the fixed effects and the imposed trends (Mora and Reggio 2012). Controlling for these trends, however, often consumes a large number of degrees of freedom and may not necessarily solve the problem if the underlying confounders are not in forms of the specified trends.

An alternative way is to model unobserved time-varying confounders semi-parametrically. For example, Bai (2009) proposes an interactive fixed effects (IFE) model, which incorporates unit-specific intercepts interacted with time-varying coefficients. The time-varying coefficients are also referred to as (latent) *factors* while the unit-specific intercepts are labelled as *factor loadings*. This approach builds upon an earlier literature on factor

[1]See Hsiao, Ching and Wan (2012) and Angrist, Jord and Kuersteiner (2013) for alternative matching methods along this line of thought.

[2]To gauge the uncertainty of the estimated treatment effect, the synthetic control method compares the estimated treatment effect with the "effects" estimated from placebo tests in which the treatment is randomly assigned to a control unit.

models in quantitative fiance.[3] The model is estimated by iteratively conducting a factor analysis of the residuals from a linear model and estimating the linear model that takes into account the influences of a fixed number of most influential factors. Pang (2010, 2014) explores non-linear IFE models with exogenous covariates in a Bayesian multi-level framework. Stewart (2014) provides a general framework of estimating IFE models based on a Bayesian variational inference algorithm. Gobillon and Magnac (2013) show that IFE models out-perform the synthetic control method in DID settings when factor loadings of the treatment and control groups do not share common support.[4]

This paper proposes a *generalized synthetic control (GSC)* method that links the two approaches and unifies the synthetic control method with linear fixed effects models under a simple framework, of which DID is a special case. It first estimates an IFE model using only the control group data, obtaining a fixed number of latent factors. It then estimates factor loadings for each treated unit by linearly projecting pre-treatment treated outcomes onto the space spanned by these factors. Finally, it imputes treated counterfactuals based on the estimated factors and factor loadings. The main contribution of this paper, hence, is to employ a latent factor approach to address a causal inference problem and provide valid uncertainty estimates under reasonable assumptions.

This method is in the spirit of the synthetic control method in the sense that by essence it is a reweighting scheme that takes pre-treatment treated outcomes as benchmarks when choosing weights for control units and uses cross-sectional correlations between treated and control units to predict treated counterfactuals. Unlike the synthetic matching method, however, it conducts dimension reduction prior to reweighting such that vectors to be reweighted on are smoothed across control units. The method can also be understood as a bias correction procedure for IFE models when the treatment effect is

[3]See Campbell, Lo and MacKinlay (1997) for applications of factor models in finance.

[4]For more empirical applications of the IFE estimator, see Kim and Oka (2014) and Gaibulloev, Sandler and Sul (2014).

heterogeneous across units.[5] It treats counterfactuals of treated units as missing data and makes out-of-sample predictions for post-treatment treated outcomes based on an IFE model.

This method has several advantages. First, it generalizes the synthetic control method to cases of multiple treated units and/or variable treatment periods. Since the IFE model is estimated only once, treated counterfactuals are obtained in a single run. Users therefore no longer need to find matches of control units for each treated unit one by one.[6] This makes the algorithm fast and less sensitive to the idiosyncrasies of a small number of observations.

Second, the GSC method produces normal frequentist uncertainty estimates, such as standard errors and confidence intervals, and improves efficiency under correct model specifications. A parametric bootstrap procedure based on simulated treated counterfactuals can provide valid inference under reasonable assumptions. Since no observations are discarded from the control group, this method uses more information from the control group and thus is more efficient than the synthetic matching method when the model is correctly specified.

Third, it embeds a cross-validation scheme that selects the number of factors of the IFE model automatically, and thus is easy to implement. One advantage of the DID data structure is that treated observations in pre-treatment periods can naturally serve as a validation dataset for model selection. I show that with sufficient data, the cross-validation procedure can pick up the correct number of factors with high probability, therefore reducing the risks of over-fitting.

[5]When the treatment effect is heterogeneous (as it is almost always the case), an IFE model that imposes a constant treatment effect assumption gives biased estimates of the average treatment effect because the estimation of the factor space is affected by the heterogeneity in the treatment effect.

[6]For examples, Acemoglu et al. (2013), who estimate the effect of Tim Geithner connections on stock market returns, conduct the synthetic control method repeatedly for each connected (treated) firm; Dube and Zipperer (2015) estimate the effect of minimum wage policies on wage and employment by conducting the method for each of the 29 policy changes. The latter also extend Abadie, Diamond and Hainmueller (2010)'s original inferential method to the case of multiple treated units using the mean percentile ranks of the estimated effects.

The GSC method has two main limitations. First, it requires more pre-treatment data than fixed effects estimators. When the number of pre-treatment periods is small, "incidental parameters" can lead to biased estimates of the treatment effects. Second, and perhaps more importantly, modelling assumptions play a heavier role with the GSC method than the original synthetic matching method. For example, if the treated and control units do not share common support in factor loadings, the synthetic matching method may simply fail to construct a synthetic control unit. Since such a problem is obvious to users, the chances that users misuse the method are small. The GSC method, however, will still impute treated counterfactuals based on model extrapolation, which may lead to erroneous conclusions. To safeguard against this risk, diagnostic checks, such as plotting the raw data and fitted values, are crucial.

## Framework

Suppose $Y_{it}$ is the outcome of interest of unit $i$ at time $t$. Let $\mathcal{T}$ and $\mathcal{C}$ denote the sets of units in treatment and control groups, respectively. The total number of units is $N = N_{tr} + N_{co}$, where $N_{tr}$ and $N_{co}$ are the numbers of treated and control units, respectively. All units are observed for $T$ periods (from time 1 to time $T$). Let $T_{0,i}$ be the number of pre-treatment periods for unit $i$, which is first exposed to the treatment at time $(T_{0,i} + 1)$ and subsequently observed for $q_i = T - T_{0,i}$ periods. Units in the control group are never exposed to the treatment in the observed time span. For notational convenience, I assume that all treated units are first exposed to the treatment at the same time, i.e., $T_{0,i} = T_0$ and $q_i = q$; variable treatment periods can be easily accommodated. First, we assume that $Y_{it}$ is given by a linear factor model.

**Assumption 1** *Functional form:*

$$Y_{it} = \delta_{it} D_{it} + x'_{it}\beta + \lambda'_i f_t + \varepsilon_{it},$$

where the treatment indicator $D_{it}$ equals 1 if unit $i$ has been exposed to the treatment prior to time $t$ and equals 0 otherwise (i.e., $D_{it} = 1$ when $i \in \mathcal{T}$ and $t > T_0$ and $D_{it} = 0$ otherwise).[7] $\delta_{it}$ is the heterogeneous treatment effect on unit $i$ at time $t$; $x_{it}$ is a $(k \times 1)$ vector of observed covariates, $\beta = [\beta_1, \cdots, \beta_k]'$ is a $(k \times 1)$ vector of unknown parameters,[8] $f_t = [f_{1t}, \cdots, f_{rt}]'$ is an $(r \times 1)$ vector of unobserved common factors, $\lambda_i = [\lambda_{i1}, \cdots, \lambda_{ir}]'$ is an $(r \times 1)$ vector of unknown factor loadings, and $\varepsilon_{it}$ represents unobserved idiosyncratic shocks for unit $i$ at time $t$ and has zero mean. Assumption 1 requires that the treated and control units are affected by the same set of factors and the number of factors is fixed during the observed time periods, i.e., no structural breaks are allowed.

The factor component of the model, $\lambda_i' f_t = \lambda_{i1} f_{1t} + \lambda_{i2} f_{2t} + \cdots + \lambda_{ir} f_{rt}$, takes a linear, additive form by assumption. In spite of the seemingly restrictive form, it covers a wide range of unobserved heterogeneities. First and foremost, conventional additive unit and time fixed effects are special cases. To see this, if we set $f_{1t} = 1$ and $\lambda_{i2} = 1$ and rewrite $\lambda_{i1} = \alpha_i$ and $f_{2t} = \xi_t$, then $\lambda_{i1} f_{1t} + \lambda_{i2} f_{2t} = \alpha_i + \xi_t$.[9] Moreover, the term also incorporates cases ranging from unit-specific linear or quadratic time trends to autoregressive components that researchers often control for when analyzing TSCS data.[10] In general, as long as an unobserved random variable can be decomposed into a multiplicative form, i.e., $U_{it} = a_i \times b_t$, it can be absorbed by $\lambda_i' f_t$ while it cannot capture unobserved confounders that are independent across units.

To formalize the notion of causality, I also use the notation from the potential outcomes

[7] Cases in which the treatment switches on and off (or "multiple-treatment-time") can be easily incorporated in this framework as long as we impose assumptions on how the treatment affects current and future outcomes. For example, one can assume that the treatment only affect the current outcome but not future outcomes (no carryover effect), as fixed effects models often do. In this paper, I do not impose such assumptions. See Imai and Kim (2016) for a thorough discussion.

[8] $\beta$ is assumed to be constant across space and time mainly for the purpose of fast computation in the frequentist framework. It is a limitation compared with more flexible and increasingly popular random coefficient models in Bayesian multi-level analysis.

[9] For this reason, additive unit and time fixed effects are not explicitly assumed in the model. A extended model that directly imposes additive two-way fixed effects is discussed in the next section.

[10] In the former case, we can set $f_{1t} = t$ and $f_{2t} = t^2$; in the latter case, for example, we can rewrite $Y_{it} = \rho Y_{i,t-1} + x_{it}'\beta + \varepsilon_{it}$ as $Y_{it} = Y_{i0} \cdot \rho^t + x_{it}'\beta + \nu_{it}$, in which $\nu_{it}$ is an AR(1) process and $\rho^t$ and $Y_{i0}$ are the unknown factor and factor loadings, respectively. See Gobillon and Magnac (2013) for more examples.

framework for causal inference (Neyman 1923; Rubin 1974; Holland 1986). Let $Y_{it}(1)$ and $Y_{it}(0)$ be the potential outcomes for individual $i$ at time $t$ when $D_{it}=1$ or $D_{it}=0$, respectively. We thus have $Y_{it}(0)=x'_{it}\beta+\lambda'_i f_t+\varepsilon_{it}$ and $Y_{it}(1)=\delta_{it}+x'_{it}\beta+\lambda'_i f_t+\varepsilon_{it}$. The individual treatment effect on treated unit $i$ at time $t$ is therefore $\delta_{it}=Y_{it}(1)-Y_{it}(0)$ for any $i\in\mathcal{T}, t>T_0$.

We can rewrite the DGP of each unit as:

$$Y_i = D_i \circ \delta_i + X_i\beta + F\lambda_i + \varepsilon_i, \quad i\in 1,2,\cdots N_{co}, N_{co}+1,\cdots,N,$$

where $Y_i=[Y_{i1},Y_{i2},\cdots,Y_{iT}]'$; $D_i=[D_{i1},D_{i2},\cdots,D_{iT}]'$ and $\delta_i=[\delta_{i1},\delta_{i2},\cdots,\delta_{iT}]'$ (symbol "$\circ$" stands for point-wise product); $\varepsilon_i=[\varepsilon_{i1},\varepsilon_{i2},\cdots,\varepsilon_{iT}]'$ are $(T\times 1)$ vectors; $X_i=[x_{i1},x_{i2},\cdots,x_{iT}]'$ is a $(T\times k)$ matrix; and $F=[f_1,f_2,\cdots,f_T]'$ is a $(T\times r)$ matrix.

The control and treated units are subscripted from 1 to $N_{co}$ and from $N_{co}+1$ to $N$, respectively. The DGP of a control unit can be expressed as: $Y_i=X_i\beta+F\lambda_i+\varepsilon_i,\ i\in 1,2,\cdots N_{co}$. Stacking all control units together, we have:

$$Y_{co}=X_{co}\beta+F\Lambda'_{co}+\varepsilon_{co}, \tag{1}$$

in which $Y_{co}=[Y_1,Y_2,\cdots,Y_{N_{co}}]$ and $\varepsilon_{co}=[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{N_{co}}]$ are $(T\times N_{co})$ matrices; $X_{co}$ is a three dimensional $(T\times N_{co}\times p)$ matrix; and $\Lambda_{co}=[\lambda_1,\lambda_2,\cdots,\lambda_{N_{co}}]'$ is a $(N_{co}\times r)$ matrix, hence, the products $X_{co}\beta$ and $F\Lambda'_{co}$ are also $(T\times N_{co})$ matrices. To identify $\beta$, $F$ and $\Lambda_{co}$ in Equation (1), more constraints are needed. Following Bai (2003, 2009), I add two sets of constraints on the factors and factor loadings: (1) all factor are normalized, and (2) they are orthogonal to each other, i.e.:

$$F'F/T=I_r \text{ and } \Lambda'_{co}\Lambda_{co}=\text{diagonal.}$$ [11]

[11]These constraints do not lead to loss of generality because for an arbitrary pair of matrices $F$ and $\Lambda_{co}$, we can find an $(r\times r)$ invertible matrix $A$ such that $(FA)'(FA)/T=I_r$ and $(A^{-1}\Lambda_{co})'A^{-1}\Lambda_{co}$ is a diagonal matrix. To see this, we can then rewrite $\lambda'_i F$ as $\tilde{\lambda}'_i\tilde{F}$, in which $\tilde{F}=FA$ and $\tilde{\lambda}_i=A^{-1}\lambda_i$ for units in both the treatment and control groups such that $\tilde{F}$ and $\tilde{\Lambda}_{co}$ satisfy the above constraints. The total number of constraints is $r^2$, the dimension of the matrix space where $A$ belongs. It is worth noting that although the original factors $F$ may not be identifiable, the space spanned by $F$, a $r$-dimensional subspace of in the $T$-dimensional space, is identified under the above constraints because for any vector in the subspace spanned by $\tilde{F}$, it is also in the subspace spanned by the original factors $F$.

For the moment, the number of factors $r$ is assumed to be known. In the next section, I propose a cross-validation procedure that automates the choice of $r$.

The main quantity of interest of this paper is the average treatment effect on the treated (ATT) at time $t$ (when $t > T_0$):

$$ATT_{t,t>T_0} = \mathbb{E}[Y_{it}(1) - Y_{it}(0)|D_{it} = 1] = \mathbb{E}[\delta_{it}|D_{it} = 1].$$ [12]

Because $Y_{it}(1)$ is observed for treated units in post-treatment periods, the main objective of this paper is to construct counterfactuals for each treated unit in post-treatment periods, i.e., $Y_{it}(0)$ for $i \in \mathcal{T}$ and $t > T_0$. The problem of causal inference indeed turns into a problem of forecasting missing data.[13]

**Assumptions for causal identification.** In addition to the functional form assumption (Assumption 1), three assumptions are required for the identification of the quantities of interest. Among them, the assumption of strict exogeneity is the most important.

**Assumption 2** *Strict exogeneity.*

$$\varepsilon_{it} \perp\!\!\!\perp D_{js}, x_{js}, \lambda_j, f_s, \quad \forall i, j, t, s.$$

Assumption 2 means that the error term of any unit at any time period is independent of treatment assignment, observed covariates, and unobserved cross-sectional and temporal heterogeneities of all units (including itself) at all periods. We call it a *strict* exogeneity assumption. It implies that treatment assignment is ignorable to potential outcomes after we condition on observed covariates and $r$ orthogonal, unobserved latent factors, i.e.,

$$\{Y_{it}(1), Y_{it}(0)\} \perp\!\!\!\perp D_{it}|x_{it}, \lambda_i, f_t, \quad \forall i, t.$$

[12]For a clear and detailed explanation of quantities of interest in TSCS analysis, see Blackwell and Glynn (2015). Using their terminology, this paper intends to estimate the Average Treatment History Effect on the Treated given two specific treatment histories: $\mathbb{E}[Y_{it}(\underline{a}_t^1) - Y_{it}(\underline{a}_t^0)|\underline{D}_{i,t-1} = \underline{a}_{t-1}^1]$ in which $\underline{a}_t^0 = (0, \cdots, 0)$, $\underline{a}_t^1 = (0, \cdots, 0, 1, \cdots, 1)$ with $T_0$ zeros and $(t - T_0)$ ones indicate the histories of treatment statuses. I keep the current notation for simplicity.

[13]The idea of predicting treated counterfactuals in a DID setup is also explored by Brodersen et al. (2014) using a structural Bayesian time series approach.

and conditional mean independence, i.e., $\mathbb{E}[\varepsilon_{it}|D_{it}, x_{it}, \lambda_i, f_t] = \mathbb{E}[\varepsilon_{it}|x_{it}, \lambda_i, f_t] = 0$. Note that because $\varepsilon_{it}$ is independent of $D_{is}$ and $x_{is}$ for all $(t, s)$, Assumption 2 rules out the possibility that past outcomes may affect future treatments, which is allowed by the so called "sequential exogeneity" assumption.[14]

Assumption 2 is arguably weaker than the strict exogeneity assumption required by fixed effects models when decomposable time-varying confounders are at present. These confounders are decomposable if they can take forms of heterogeneous impacts of a common trend or a series of common shocks. For instance, suppose a law is passed in a state because the public opinion in that state becomes more liberal. Because changing ideologies are often cross-sectionally correlated across states, a latent factor may be able to capture shifting ideology at the national level; the national shifts may have a larger impact on a state that has a tradition of mass liberalism or has a higher proportion of manufacturing workers than a state that is historically conservative. Controlling for this unobserved confounder, therefore, can alleviate the concern that the passage of the law is endogenous to changing ideology of a state's constituents to a great extent.

When such a confounder exists, with two-ways fixed effects models we need to assume that $(\varepsilon_{it} + \lambda_i f_t) \perp\!\!\!\perp D_{js}, x_{js}, \alpha_j, \xi_s, \forall i, j, t, s$ (with $\lambda_i f_t$, $\alpha_j$ and $\xi_s$ representing the time-varying confounder for unit $i$ at time $t$, fixed effect for unit $j$, and fixed effect for time $s$, respectively) for the identification of the constant treatment effect. This is implausible because $\lambda_i f_t$ is likely to be correlated with $D_{it}$, $x_{it}$, and $\alpha_i$, not to mention other terms. In contrast, Assumption 2 allows the treatment indicator to be correlated with both $x_{js}$ and $\lambda_j' f_s$ for any unit $j$ at any time periods $s$ (including $i$ and $t$ themselves).

Identifying the treatment effects also requires the following assumptions.

**Assumption 3** *Weak serial dependence of the error terms.*

[14] A directed acyclic graph (DAG) representation is provided in the Appendix (Figure 1). See Blackwell and Glynn (2015) and Imai and Kim (2016) for discussions on the difference between the strict ignorability and sequential ignorability assumptions. What is unique here is that we conditional on unobserved factors and factor loadings.

**Assumption 4** *Regularity conditions.*

Assumptions 3 and 4 (see Appendix for details) are needed for the consistent estimation of $\beta$ and the space spanned by $F$ (or $F'F/T$). Similar, though slightly weaker, assumptions are made in Bai (2009) and Moon and Weidner (2013). Assumption 3 allows weak serial correlations but rules out strong serial dependence, such as unit root processes; errors of different units are uncorrelated. A sufficient condition for Assumption 3 to hold is that the error terms are not only independent of covariates, factors and factor loadings, but also independent both across units and over time, which is assumed in Abadie, Diamond and Haimmueller (2010). Assumption 4 specifies moment conditions that ensure the convergence of the estimator.

For valid inference based on a block bootstrap procedure discussed in the next section, we also need to Assumption 5 (see Appendix for details). Heteroskedasticity across time, however, is allowed.

**Assumption 5** *The error terms are cross-sectionally independent and homoscedastic.*

**Remark 1:** Assumptions 3 and 5 suggest that the error terms $\varepsilon_{it}$ can be serially correlated. Assumption 2 rules out dynamic models with lagged dependent variables, however, this is mainly for the purpose of simplifying proofs (Bai 2009, p. 1243). As long as the error terms are not serially correlated, the propose method can accommodate dynamic models.

## Estimation Strategy

In this section, I first propose a generalized synthetic control (GSC) estimator for treatment effect of each treated unit. It is essentially an out-of-sample prediction method based on Bai (2009)'s factor augmented model.

The GSC estimator for the treatment effect on treated unit $i$ at time $t$ is given by the difference between the actual outcome and its estimated counterfactual: $\hat{\delta}_{it} = Y_{it}(1) -$

$\hat{Y}_{it}(0)$, in which $\hat{Y}_{it}(0)$ is imputed with three steps. In the first step, we estimate an IFE model using only the control group data and obtain $\hat{\beta}, \hat{F}, \hat{\Lambda}_{co}$:

$$\textbf{Step 1.} \qquad \left(\hat{\beta}, \hat{F}, \hat{\Lambda}_{co}\right) = \underset{\tilde{\beta}, \tilde{F}, \tilde{\Lambda}_{co}}{\text{argmin}} \sum_{i \in \mathcal{C}} (Y_i - X_i\tilde{\beta} - \tilde{F}\tilde{\lambda}_i)'(Y_i - X_i\tilde{\beta} - \tilde{F}\tilde{\lambda}_i)$$

$$s.t. \quad \tilde{F}'\tilde{F}/T = I_r \text{ and } \tilde{\Lambda}_{co}'\tilde{\Lambda}_{co} = \text{diagonal}.$$

Appendix C explains in detail how such a model is estimated. The second step estimates factor loadings for each treated unit by minimizing the mean squared error of the predicted treated outcome in pre-treatment periods:

$$\begin{aligned} \textbf{Step 2.} \qquad \hat{\lambda}_i &= \underset{\tilde{\lambda}_i}{\text{argmin}} (Y_i^0 - X_i^0\hat{\beta} - \hat{F}^0\tilde{\lambda}_i)'(Y_i^0 - X_i^0\hat{\beta} - \hat{F}^0\tilde{\lambda}_i) \\ &= (\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{F}^{0\prime}(Y_i^0 - X_i^0\hat{\beta}), \qquad i \in \mathcal{T}, \end{aligned}$$

in which $\hat{\beta}$ and $\hat{F}^0$ are from the first-step estimation and the superscripts "0"s denote the pre-treatment periods. In the third step, we calculate treated counterfactuals based on $\hat{\beta}$, $\hat{F}$, and $\hat{\lambda}_i$:

$$\textbf{Step 3.} \qquad \hat{Y}_{it}(0) = x_{it}'\hat{\beta} + \hat{\lambda}_i'\hat{f}_t \qquad i \in \mathcal{T},\ t > T_0.$$

An estimator for $ATT_t$ therefore is: $\widehat{ATT}_t = \frac{1}{N_{tr}} \sum_{i \in \mathcal{T}} [Y_{it}(1) - \hat{Y}_{it}(0)]$ for $t > T_0$.

**Remark 2:** In Appendix E, we show that, under Assumption 1-4, the bias of the GSC shrinks to zero as the sample size grows, i.e. $\mathbb{E}_\varepsilon(\widehat{ATT}_t \mid D, X, \Lambda, F) \to ATT_t$ as $N_{co}, T_0 \to 0$ ($N_{tr}$ is taken as given).[15] Intuitively, both large $N_{co}$ and large $T_0$ are necessary for the convergences of $\hat{\beta}$ and the estimated factor space. When $T_0$ is small, imprecise estimation of the factor loadings, or the "incidental parameters" problem, will lead to bias in the estimated treatment effects. This is a crucial difference from the conventional linear

[15] $D = [D_1, D_2, \cdots, D_N]$ is a $(T \times N)$ matrix, $X$ is a three dimensional $(T \times N \times p)$ matrix; and $\Lambda = [\lambda_1, \lambda_2, \cdots, \lambda_N]'$ is a $(N \times r)$ matrix.

fixed-effect models.

**Model selection.** In practice, researchers may have limited knowledge of the exact number of factors to be included in the model. Therefore, I develop a cross-validation procedure to select models before estimating the causal effect. It relies on the control group information as well as information from the treatment group in pre-treatment periods. Algorithm 1 describes the details of this procedure.

**Algorithm 1 (Cross-validating the number of factors)** *A leave-one-out-cross-validation procedure that selects the number of factors takes the following steps:*

**Step 1.** Start with a given number of factors $r$, estimate an IFE model using the control group data $\{Y_i, X_i\}_{i \in \mathcal{C}}$, obtaining $\hat{\beta}$ and $\hat{F}$;

**Step 2.** **Start** a cross-validation loop that goes through all $T_0$ pre-treatment periods:

a) In round $s \in \{1, \cdots, T_0\}$, hold back data of all treated units at time $s$. Run an OLS regression using the rest of the pre-treatment data, obtaining factor loadings for each treated unit $i$:

$$\hat{\lambda}_{i,-s} = (F^{0\prime}_{-s} F^{0}_{-s})^{-1} F^{0\prime}_{-s} (Y^{0}_{i,-s} - X^{0\prime}_{i,-s}\hat{\beta}), \qquad \forall i \in \mathcal{T},$$

in which the subscripts "-s" stands for all pre-treatment periods except for $s$.

b) Predict the treated outcomes at time $s$ using $\hat{Y}(0)_{is} = x'_{is}\hat{\beta} + \hat{\lambda}'_{i,-s}\hat{f}_s$ and save the prediction error $e_{is} = Y_{is}(0) - \hat{Y}_{is}(0)$ for all $i \in \mathcal{T}$.

**End** of the cross-validation loop;

**Step 3.** Calculate the mean square prediction error (MSPE) given $r$,

$$MSPE(r) = \sum_{s=1}^{T_0} \sum_{i \in \mathcal{T}} e^2_{is} / T_0.$$

**Step 4.** Repeat Steps 1-3 with different $r$'s and obtain corresponding MSPEs.

**Step 5.** Choose $r^*$ that minimizes the MSPE.

The basic idea of the above procedure is to hold back a small amount of data (e.g. one pre-treatment period of the treatment group) and use the rest of data to predict the held-back information. The algorithm then chooses the model that on average makes the most accurate predictions. A TSCS dataset with a DID data structure allows us to do so because (1) there exists a set of control units that are never exposed to the treatment and therefore can serve as the basis for estimating time-varying factors and (2) the pre-treatment periods of treated units constitute a natural validation set for candidate models. This procedure is computationally inexpensive because with each $r$, the IFE model is estimated only once (Step 1). Other steps involves merely simple calculations. In the next chapter, we conduct Monte Carlo exercises and show that the above procedure performs well in term of choosing the correct number of factors even with relatively small datasets.

**Remark 3:** Our framework can also accommodate DGPs that directly incorporate additive fixed effects, known time trends, and exogenous time-invariant covariates, such as:

$$Y_{it} = \delta_{it}D_{it} + x'_{it}\beta + \gamma'_i l_t + z'_i\theta_t + \lambda'_i f_t + \alpha_i + \xi_t + \varepsilon_{it}, \tag{2}$$

in which $l_t$ is a $(q \times 1)$ vector of known time trends that may affect each unit differently; $\gamma_i$ is $(q \times 1)$ unit-specific unknown parameters; $z_i$ is a $(m \times 1)$ vector of observed time-invariant covariates; $\theta_t$ is a $(m \times 1)$ vector of unknown parameters; $\alpha_i$ and $\xi_t$ are additive individual and time fixed effects, respectively.[16] Appendix D describes the estimation procedure of this extended model.

[16] As mentioned in Section , Equation (2) can be represented by the equation specified in Assumption 1 when we set $\lambda_i^l = \gamma_i$, $f_t^l = l_t$, $\lambda_i^z = z_i$, $f_t^z = \theta_t$, $\lambda_i^\alpha = \alpha_i$, $f_t^\alpha = 1$, and $\lambda_i^\xi = 1$, $f_t^\xi = \xi_t$. However, if we know that the model specified in Equation (2) is correct, explicitly including additive fixed effects and time-invariant covariates in the model improves efficiency. Such a model, although requiring more data, relies on an identification assumption that is arguably more appealing than Assumption 2 since time-invariant heterogeneities, universal shocks over time, differential impacts of known time trends, and differential trends caused by observed time-invariant covariates, are all being explicitly conditioned on, i.e., $\{Y_{it}(1), Y_{it}(0)\} \perp\!\!\!\perp D_{it}|x_{it}, l_t, z_i, \alpha_i, \xi_t, \lambda_i, f_t \quad \forall i, t$.

## Inference

I rely on a parametric bootstrap procedure to obtain the uncertainty estimates of the GSC estimator.[17] When the sample size is large, when $N_{tr}$ is large in particular, a simple non-parametric bootstrap procedure can provide valid uncertainty estimates. When the sample size is small, especially when $N_{tr}$ is small, we are unable to approximate the DGP of the treatment group by resampling the data non-parametrically. In this case, we simply lack the information of the joint distribution of $(X_i, \lambda_i, \delta_i)$ for the treatment group. However, we can obtain uncertainty estimates conditional on observed covariates and unobserved factors and factor loadings using a parametric bootstrap procedure via re-sampling the errors.[18] Our goals is to estimate the conditional variance of ATT estimator, i.e.,

$$\text{Var}_\varepsilon(\widehat{ATT}_t \mid D, X, \Lambda, F) = \text{Var}_\varepsilon\left(\frac{1}{N_{tr}} \sum_{i \in \mathcal{T}} \{\varepsilon_{it} - x'_{it}\hat{\beta} - \hat{\lambda}'_i \hat{f}_t\} \mid D, X, \Lambda, F\right).$$

Since $(D_i, X_i, \lambda_i, \delta_i)$ are taken as given, the only remaining random variable that is not being conditioned on is $\varepsilon_i$, which are assumed to be independent of treatment assignment, observed covariates, factors and factor loadings (Assumption 2). We can interpret $\varepsilon_i$ as measurement errors or sources of variations in the outcome that we cannot explain but are unrelated to treatment assignment.[19]

In the parametric bootstrap procedure, we simulate treated counterfactuals and control units based on the following re-sampling scheme:

$$\tilde{Y}_i(0) = X_i\hat{\beta} + \hat{F}\hat{\lambda}_i + \tilde{\varepsilon}_i, \qquad \forall i \in \mathcal{C};$$
$$\tilde{Y}_i(0) = X_i\hat{\beta} + \hat{F}\hat{\lambda}_i + \tilde{\varepsilon}^p_i, \qquad \forall i \in \mathcal{T}.$$

[17] Under certain restrictive conditions, such as independent, identically and normally distributed errors, it is possible to derive the analytical asymptotic distribution of the GSC estimator, a necessary step for future research.
[18] By re-sampling entire time-series of error terms, we preserve the serial correlation within the units, thus avoiding underestimating the standard errors due to serial correlations (Beck and Katz 1995).
[19] $\varepsilon_{it}$ may be correlated with $\hat{\lambda}_i$ when the errors are serially correlated because $\hat{\lambda}_i$ is estimated using the pre-treatment data.

in which $\tilde{Y}_i(0)$ is a vector of simulated treated counterfactuals or control outcomes; $X_i\hat{\beta} + \hat{F}\hat{\lambda}_i$ is the estimated conditional mean; and $\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_i^p$ are re-sampled errors for unit $i$, depending on whether it belongs to the treatment or control group. $\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_i^p$ are drawn from different empirical distributions because $\hat{\beta}$ are $\hat{F}$ are estimated using only the control group information; hence, $X_i\hat{\beta} + \hat{F}\hat{\lambda}_i$ predicts $X_i\beta + F\lambda_i$ better for a control unit than for a treated unit (as a result, the variance of $\tilde{\varepsilon}_i^p$ is usually bigger than that of $\tilde{\varepsilon}_i$). $\tilde{\varepsilon}_i$ is the in-sample error of the IFE model fitted to the control group data, and therefore is drawn from the empirical distribution of the residuals of the IFE model, while $\tilde{\varepsilon}_i^p$ can be seen as the prediction error of the IFE model for treated counterfactuals.[20]

Although we cannot observe treated counterfactuals, $Y_{it}(0)$ is observed for all control units. With the assumptions that treated and control units follow the same factor model (Assumption 1) and the error terms are independent and homoscedastic across space (Assumption 5), we can use a cross-validation method to simulate $\varepsilon_i^p$ based on the control group data (Efron 2012). Specifically, each time we leave one control unit out (to be taken as a "fake" treat unit) and use the rest of the control units to predict the outcome of left-out unit. The difference between the predicted and observed outcomes is a prediction error of the IFE model. $\varepsilon_i^p$ is drawn from the empirical distributions of the prediction errors. Under Assumptions 1-5, this procedure provides valid uncertainty estimates for the proposed method without making particular distributional assumptions of the error terms. Algorithm 2 describes the entire procedure in detail.

[20]The treated outcome for unit $i$, thus can be drawn from $\tilde{Y}_i(1) = \tilde{Y}_i(0) + \delta_i$. We do not directly observe $\delta_i$, but since it is taken as given (a set of fixed numbers), its presence will not affect the uncertainty estimates of $\widehat{ATT}_t$. Hence, in the bootstrap procedure, I use $\tilde{Y}_i(0)$ for both the treatment and control groups to form bootstrapped samples (set $\delta_i = \mathbf{0}$, for all $i \in \mathcal{T}$). We will add back $\widehat{ATT}_t$ when constructing confidence intervals.

**Algorithm 2 (Inference)** *A parametric bootstrap procedure that gives the uncertainty estimates of the ATT is described as follows:*

**Step 1. Start** a loop that runs $B_1$ times:

a) In round $m \in \{1, \cdots, B_1\}$, randomly select one control unit $i$ as if it was treated when $t > T_0$;

b) Re-sample the rest of control group with replacement of size $N_{co}$ and form a new sample with one "treated" unit and $N_{co}$ re-sampled control units;

c) Apply the GSC method to the new sample, obtaining a vector of prediction error; $\hat{\varepsilon}^p_{(m)} = Y_i - \hat{Y}_i(0)$.

**End** of the loop, collecting $\mathbf{\hat{e}^p} = \{\hat{\varepsilon}^p_{(1)}, \hat{\varepsilon}^p_{(2)} \cdots, \hat{\varepsilon}^p_{(B_1)}\}$.

**Step 2.** Apply the GSC method to the original data, obtaining: (1) $\widehat{ATT}_t$ for all $t > T_0$, (2) estimated coefficients: $\hat{\beta}$, $\hat{F}$, $\hat{\Lambda}_{co}$, and $\hat{\lambda}_{j,j\in\mathcal{T}}$, and (3) the fitted values and residuals of the control units: $\mathbf{\hat{Y}_{co}} = \{\hat{Y}_1(0), \hat{Y}_2(0), \cdots, \hat{Y}_{N_{co}}(0)\}$ and $\mathbf{\hat{e}} = \{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_{N_{co}}\}$.

**Step 3. Start** a bootstrap loop that runs $B_2$ times:

a) In round $k \in \{1, \cdots, B_2\}$, construct a bootstrapped sample $S^{(k)}$ by:

$$\begin{aligned} \tilde{Y}_i^{(k)}(0) &= \hat{Y}_i(0) + \tilde{\varepsilon}_i, \quad i \in \mathcal{C} \\ \tilde{Y}_i^{(k)}(0) &= \hat{Y}_i(0) + \tilde{\varepsilon}_i^p, \quad j \in \mathcal{T} \end{aligned}$$

in which each vector of $\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_j^p$ are randomly selected from sets $\mathbf{e}$ and $\mathbf{e}^p$, respectively, and $\hat{Y}_i(0) = X_i\hat{\beta} + \hat{F}\hat{\lambda}_i$. Note that the simulated treated counterfactuals do not contain the treatment effect.

b) Apply the GSC method to $S^{(k)}$ and obtain a new ATT estimate; add $\widehat{ATT}_{t,t>T_0}$ to it, obtaining the bootstrapped estimate $\widehat{ATT}^{(k)}_{t,t>T_0}$.

**End** of the bootstrap loop.

**Step 4.** Compute the variance of $\widehat{ATT}_{t,t>T_0}$ using

$$\mathrm{Var}(\widehat{ATT}_t \mid D, X, \Lambda, F) = \tfrac{1}{B}\sum_{k=1}^{B}\left(\widehat{ATT}_t^{(k)} - \tfrac{1}{B}\sum_{j=1}^{B}\widehat{ATT}_t^{(j)}\right)^2$$

and its confidence interval using the conventional percentile method (Efron and Tibshirani 1993).

## Conclusion

In this chapter, I propose the generalized synthetic control (GSC) method for causal inference with TSCS data. It attempts to address the challenge that the "parallel trends" assumption often fails when researchers apply fixed effects models to estimate the causal effect of a certain treatment. The GSC method estimates the individual treatment effect on each treated unit semi-parametrically. Specifically, it imputes treated counterfactuals based on a linear interactive fixed effects model that incorporates time-varying coefficients (factors) interacted with unit-specific intercepts (factor loadings). A built-in cross-validation scheme automatically selects the model, reducing the risks of over-fitting.

This method is in spirit of the original synthetic control method in that it uses data from pre-treatment periods as benchmarks to customize a re-weighting scheme of control units in order to make the best possible predictions for treated counterfactuals. It generalizes the synthetic control method in two aspects. First, it allows multiple treated units and differential treatment timing. Second, it offers uncertainty estimates, such as standard errors and confidence intervals, that are easy to interpret.

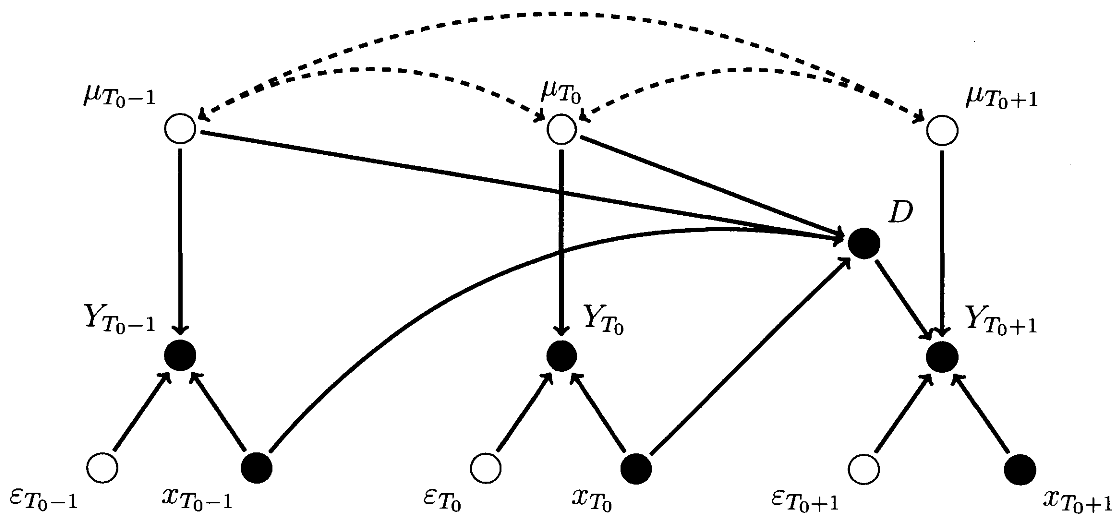
Two caveats are worth emphasizing when applying this method. First, insufficient data (with either a small $T_0$ or a small $N_{co}$) cause bias in the estimation of the treatment effect.[21] Second, excessive extrapolations based on imprecisely estimated factors and factor loading can lead to erroneous results. To avoid this problem, I recommend the following diagnostics upon using this method: (1) plot raw data of treated and control outcomes as well as imputed counterfactuals and check whether the imputed values are within reasonable intervals; (2) plot estimated factor loadings of both treated and control units and check the overlap.[22] When excessive extrapolations appear to happen, we recommend users to include a smaller number of factors or switch back to the conventional DID framework.

[21] Users should be cautious about using this method when $T_0 < 10$ and $N_{co} < 40$.
[22] We provide software routines that can generate these diagnostic plots automatically.

# Appendix: Technical Details

## A. A Directed Acyclic Graph (DAG)

Figure 1.. A DAG Illustration




**Note:** Unit indices are dropped for simplicity. Vector $\mu_t$ represents unobserved time-varying confounders. If Assumption 1 holds, $\mu_t$ (or $\mu_{it}$) can be expressed as $\lambda_i' f_t$. We allow $D_i$ to be correlated with $x_{is,s<t}$ and $\mu_{is,s<t}$. In fact, we also allow it to be correlated with $x_{js,s<t}$ and $\mu_{js,s<t}$ when $j \neq i$.

## B. Technical Assumptions

Assumptions 3–5 are shown below.

**Assumption 3** Weak serial dependence of the error terms:

1. $\mathbb{E}(\varepsilon_{it}\varepsilon_{is}) = \sigma_{i,ts}$, $|\sigma_{i,ts}| \leq \bar{\sigma}_i$ for all $(t,s)$ such that $\frac{1}{N}\sum_{i=1}^{N}\bar{\sigma}_i \leq M$.
2. For every $(t,s)$, $\mathbb{E}|N^{-1/2}\sum_{i=1}^{N}[\varepsilon_{is}\varepsilon_{it} - \mathbb{E}(\varepsilon_{is}\varepsilon_{it})]|^4 \leq M$.
3. $\frac{1}{T^2N}\sum_{t,s,u,v}\sum_{i,j}|cov(\varepsilon_{it}\varepsilon_{is}, \varepsilon_{ju}\varepsilon_{jv})| \leq M$ and $\frac{1}{TN^2}\sum_{t,s}\sum_{i,j,k,l}|cov(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{ls})| \leq M$.
4. $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = 0$, for all $i \neq j$, $(t,s)$.

**Assumption 4** Regularity conditions:

1. $\mathbb{E}|\varepsilon_{it}|^8 \leq M$.
2. $\mathbb{E}||x_{it}||^4 \leq M$: Let $\mathcal{F} = \{F : F'F/T = I\}$. We assume $\inf_{F\in\mathcal{F}} D(F) > 0$, in which $D(F) = \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} S_i'S_i$, where $S_i = M_F X_i - \frac{1}{N_{co}}\sum_{k=1}^{N} M_F X_k a_{ik}$ and $a_{ik} = \lambda_i'(\Lambda_{co}'\Lambda_{co})^{-1}\lambda_k$.
3. $\mathbb{E}||f_t||^4 \leq M < \infty$ and $\frac{1}{T}\sum_{t=1}^{T} f_t f_t' \xrightarrow{p} \Sigma_F$ for some $r \times r$ positive definite matrix $\Sigma_F$, as $T_0 \to \infty$.
4. $\mathbb{E}||\lambda_i||^4 \leq M < \infty$ and $\Lambda_{co}'\Lambda_{co}/N_{co} \xrightarrow{p} \Sigma_N$ for some $r \times r$ positive definite matrix $\Sigma_N$, as $N_{co} \to \infty$.

**Assumption 5** The error terms are cross-sectionally independent and homoscedastic.

1. $\varepsilon_{it} \perp\!\!\!\perp \varepsilon_{js}$ for all $j \neq i$, $(t,s)$.
2. $\mathbb{E}(\varepsilon_{it}\varepsilon_{is}) = \sigma_{ts} \leq M$, for all $(t,s)$.

## C. Estimating an Interactive Fixed-effect Model

As in Equation (1), I assume that the control units follow an interactive fixed-effect model:

$$Y_{co} = X_{co}\beta + F\Lambda_{co}' + \varepsilon_{co},$$

The least square objective function is

$$SSR(\beta, F, \Lambda_{co}) = \sum_{i=1}^{N_{co}} (Y_i - X_i\beta - F\lambda_i)'(Y_i - X_i\beta - F\lambda_i).$$

The goal is to estimate $\beta$, $F$, and $\Lambda_{co}$ by minimizing the SSR subject to the following constraints:

$$F'F/T = I_r \text{ and } \Lambda_{co}'\Lambda_{co} = \text{diagonal}.$$

A unique solution $(\hat{\beta}, \hat{F}, \hat{\Lambda}_{co})$ to this problem exists. To find the solution, Bai (2009) proposed an iteration scheme that can lead to the unique solution starting from some initial value of $\beta$ (for instance, the least-square dummy-variable (LSDV) estimates) or $F$. In each iteration, given $F$ and $\Lambda_{co}$, the algorithm computes $\beta$:

$$\hat{\beta}(F, \Lambda) = \left( \sum_{i=1}^{N_{co}} X_i'X_i \right)^{-1} \sum_{i=1}^{N_{co}} X_i'(Y_i - F\lambda_i),$$

and given $\beta$, it computes $F$ and $\Lambda_{co}$ from a pure factor model $(Y_i - X_i\beta) = F\lambda_i + \varepsilon_i$:

$$\begin{aligned} [\tfrac{1}{N_{co}T} \textstyle\sum_{i=1}^{N_{co}} (Y_i - X_i\beta)(Y_i - X_i\beta)']\hat{F} &= \hat{F}V_{N_{co}T}, \\ \hat{\Lambda}_{co} &= \tfrac{1}{T}(Y - X\beta)'\hat{F}, \end{aligned}$$

in which $V_{N_{co}T}$ is a diagonal matrix that consists for the first $r$ largest eigenvalues of the $(N_{co} \times N_{co})$ matrix $\frac{1}{N_{co}T} \sum_{i=1}^{N_{co}} (Y_i - X_i\beta)(Y_i - X_i\beta)'$ and $V_{N_{co}T} = \frac{1}{N_{co}} \hat{\Lambda}_{co}'\hat{\Lambda}_{co}$.

## D. Estimation Procedure for an Extended Model

Without loss of generality, we re-write Equation (2) as

$$Y_{it} = \delta_{it} D_{it} + x'_{it}\beta + \gamma'_i l_t + z'_i \theta_t + \lambda'_i f_t + \alpha_i + \xi_t + \mu + \varepsilon_{it},$$

in which $\mu$ is the mean of control group outcomes, which allows us to impose two restrictions: $\sum_{i=1}^{N_{co}} \alpha_i = 0$ and $\sum_{i=1}^{N_{co}} \xi_i = 0$. As before, we use three steps to impute the counterfactuals for treated units. It can be written as

$$Y_i = \delta_i \circ D_i + X_i\beta + L\gamma_i + \Theta z_i + F\lambda_i + \alpha_i \mathbf{1}_T + \Xi + \mu \mathbf{1}_T + \varepsilon_i,$$

in which $L = [l_1, l_2, \cdots, l_T]'$, a $(T \times q)$ matrix; $\Theta = [\theta_1, \theta_2, \cdots, \theta_T]'$, a $(T \times m)$ matrix; and $\Xi = [\xi_1, \xi_2, \cdots, \xi_T]'$, a $(T \times 1)$ vector. In the first step, we estimate an extended IFE model using only the control group data and obtain $\hat{\beta}$,$\hat{F}$,$\hat{\Lambda}_{co}$, $\hat{\Xi}$, $\hat{\Theta}$, $\hat{\gamma}_i$, and $\hat{\alpha}_i$ (for all $i \in \mathcal{C}$) and $\hat{\mu}$:

$$\textbf{Step 1.} \quad \left(\hat{\beta}, \hat{F}, \hat{\Theta}, \hat{\Xi}, \hat{\Lambda}_{co}, \{\hat{\gamma}_i\}, \{\hat{\alpha}_i\}, \hat{\mu}\right) = \underset{\tilde{\beta},\tilde{F},\tilde{\Theta},\tilde{\Xi},\tilde{\Lambda}_{co},\{\tilde{\gamma}_i\},\{\tilde{\alpha}_i\},\tilde{\mu}}{\operatorname{argmin}} \sum_{i \in \mathcal{C}} \tilde{e}'_i \tilde{e}_i.$$

in which $\tilde{e}_i = Y_i - X_i\tilde{\beta} - L\tilde{\gamma}_i - \tilde{\Theta} z_i - \tilde{F}\tilde{\lambda}_i - \tilde{\alpha}_i \mathbf{1}_T - \tilde{\Xi} - \tilde{\mu}\mathbf{1}_T$. The details of the estimation strategy can be found in Bai (2009) Sections 8 and 10.

The second step estimates factor loadings, as well as additive unit fixed effects, for each treated unit by minimizing the mean squared error of the treated units in the pre-treatment period:

$$\begin{aligned} \textbf{Step 2.} \quad \left(\hat{\gamma}_i \ \hat{\lambda}_i \ \hat{\alpha}_i\right)' &= \underset{(\tilde{\gamma}_i \ \tilde{\lambda}_i \ \tilde{\alpha}_i)'}{\operatorname{argmin}} \underline{e}'_i \underline{e}_i \\ &= (\hat{G}^{0\prime}\hat{G}^0)^{-1}\hat{G}^{0\prime}(Y_i^0 - X_i^0\hat{\beta} - \hat{\Theta}^0 z_i - \hat{\Xi}^0 - \hat{\mu}\mathbf{1}_{T_0}), \qquad i \in \mathcal{T}. \end{aligned}$$

in which $\underline{e}_i = Y_i^0 - X_i^0\hat{\beta} - L^0\tilde{\gamma}_i - \hat{\Theta}^0 z_i - \hat{F}^0\tilde{\lambda}_i - \tilde{\alpha}_i - \hat{\xi}_t^0 - \hat{\mu}$; $\hat{\beta}$, $\hat{F}$, $\hat{\Theta}$, $\hat{\Xi}$ and $\hat{\mu}$ are from the first step estimation; the superscripts "0"s denote the pre-treatment period; and $\hat{G}^0 = (L^0 \ \hat{F}^0 \ \mathbf{1}_{T_0})$ is $(L^0 \ \hat{F}^0)$ augmented with a column of ones, a $T_0 \times (q + r + 1)$ matrix.

In the third step, we calculate the counterfactual based on $\hat{\beta}$, $\hat{F}$, $\hat{\lambda}_i$, $\hat{\alpha}_i$, $\hat{\xi}_t$, and $\hat{\mu}$:

$$\textbf{Step 3.} \quad \hat{Y}_{it}(0) = x'_{it}\hat{\beta} + \hat{\gamma}'_i l_t + z'_i\hat{\theta}_t + \hat{\lambda}'_i\hat{f}_t + \hat{\alpha}_i + \hat{\xi}_t + \hat{\mu}, \qquad \forall i \in \mathcal{T}.$$

## E. Proofs of Main Results

I use the Frobenius norm throughout this paper, i.e., for any vector or matrix $M$, its norm is defined as $\|M\| = \sqrt{\mathrm{tr}(M'M)}$. I establish four lemmas before getting to the main results.

**Lemma 3** *(i)* $T^{-1/2}\|\hat{F}^0\| = O_p(1)$; *(ii)* $T\|(\hat{F}^{0\prime}\hat{F}^0)^{-1}\| = O_p(1)$.

**Proof:** (i). Because $\mathrm{tr}(\hat{F}'\hat{F}/T) = r$,

$$T^{-1/2}\|\hat{F}^0\| = T^{-1/2}\sqrt{\mathrm{tr}(\hat{F}^{0\prime}\hat{F}^0)} \le T^{-1/2}\sqrt{\mathrm{tr}(\hat{F}'\hat{F})} = \sqrt{r}.$$

(ii). Denote $Q = \sum_{s=T_0+1}^{T} \hat{f}_s\hat{f}_s'$, a symmetric and positive definite $(r \times r)$ matrix. Because $\|\hat{f}_t\hat{f}_t'\| = O_p(1)$ and there are only $q_i$ items in the summation, $\|Q\| = O_p(1)$. Since $\hat{F}^{0\prime}\hat{F}^0 = \hat{F}'\hat{F} - Q = T\cdot I_r - Q$,

$$(\hat{F}^{0\prime}\hat{F}^0)^{-1} = \frac{1}{T}I_r + (I - \frac{1}{T}Q)^{-1}\frac{1}{T^2}Q$$

Since $Q$ is positive definite, $\|(\hat{F}^{0\prime}\hat{F}^0)^{-1}\|$ is strictly decreasing in $T$ and is $O_p(T^{-1})$.

**Lemma 4** $\|\hat{\beta} - \beta\| = O_p(N_{co}^{-1}) + O_p(T^{-1}) + o_p\left((N_{co}T)^{-1/2}\right)$.

**Proof:** Bai (2009) shows that under Assumptions 3 and 4 and when $T/N^2 \to 0$:

$$\begin{aligned}\hat{\beta} - \beta =& D(\hat{F})^{-1}\frac{1}{N_{co}T}\sum_{i=1}^{N_{co}}[X_i'M_F - \frac{1}{N_{co}}\sum_{k=1}^{N_{co}} a_{ik}X_k'M_F]\varepsilon_i \\ &+ \frac{1}{N_{co}}\xi + \frac{1}{T}\zeta + \frac{1}{\sqrt{N_{co}T}}o_p(1),\end{aligned} \tag{3}$$

$$\begin{aligned}\text{where } D(\hat{F}) =& \frac{1}{N_{co}T}\sum_{i}^{N_{co}} Z_i'Z_i, \quad Z_i = M_F X_i - \frac{1}{N_{co}}\sum_{k=1}^{N_{co}} M_F X_k a_{ik}, \\ \xi =& -D(\hat{F})^{-1}\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}\sum_{k=1}^{N_{co}}\frac{(X_i - V_i)'F}{T}\left(\frac{F'F}{T}\right)^{-1}\left(\frac{\Lambda_{co}'\Lambda_{co}}{N_{co}}\right)^{-1}\lambda_k\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\right) = O \\ \zeta =& -D(\hat{F})^{-1}\frac{1}{NT}\sum_{i=1}^{N_{co}} X_i'M_{\hat{F}}\Omega\hat{F}\left(\frac{F'\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda_{co}'\Lambda_{co}}{N_{co}}\right)^{-1}\lambda_i = O_p(1),\end{aligned}$$

and $a_{ik} = \lambda_i'(\Lambda_{co}'\Lambda_{co}/N_{co})^{-1}\lambda_k$, $V_i = \frac{1}{N_{co}}\sum_{k=1}^{N_{co}} a_{ik}X_k$, $\Omega = \frac{1}{N_{co}}\sum_{k=1}^{N_{co}} E(\varepsilon_k\varepsilon_k')$. Therefore, $\hat{\beta}$

is an asymptotically unbiased estimator for $\beta$ when both $T$ and $N_{co}$ are large and

$$\|\hat{\beta}-\beta\| = O_p(N_{co}^{-1}) + O_p(T^{-1}) + o_p\left((N_{co}T)^{-1/2}\right).$$

**Lemma 5** *Denote* $H = \left(\frac{\Lambda_{co}'\Lambda_{co}}{N_{co}}\right)\left(\frac{F'\hat{F}}{T}\right)V_{N_{co}T}^{-1}$.

*(i).* $\|f_t - H^{-1}\hat{f}_t\| = O_p(N_{co}^{-1/2}) + O_p(T^{-1/2})$;

*(ii).* $\|f_t' - \hat{f}_t{}'(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{F}^{0\prime}F^0\| = O_p(N_{co}^{-1/2}) + O_p(T^{-1/2})$.

**Proof:** (i). The main logic of this proof follows Bai (2009) Proposition A.1 (p. 1266). Because

$$\left[\frac{1}{N_{co}T}\sum_{i}^{N_{co}}(Y_i - X_i)(Y_i - X_i)'\right]\hat{F} = \hat{F}V_{N_{co}T}$$

and $Y_i - X_i\hat{\beta} = X_i(\beta-\hat{\beta}) + F\lambda_i + \varepsilon_i$, by expanding the terms on the left-hand side, we have:

$$\begin{aligned}
\hat{F}V_{N_{co}T} =& \frac{1}{N_{co}T}\sum_{i}^{N_{co}} X_i(\beta-\hat{\beta})(\beta-\hat{\beta})'X_i'\hat{F} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} X_i(\beta-\hat{\beta})\lambda_i'F'\hat{F} \\
&+ \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} X_i(\beta-\hat{\beta})\varepsilon'\hat{F} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} F\lambda_i(\beta-\hat{\beta})'X_i'\hat{F} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} \varepsilon_i(\beta-\hat{\beta})'X_i'\hat{F} \\
&+ \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} F\lambda_i\varepsilon_i'\hat{F} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} \varepsilon_i\lambda_i'F'\hat{F} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} \varepsilon_i\varepsilon_i'\hat{F} + F\left(\frac{\Lambda_{co}'\Lambda_{co}}{N_{co}}\right)\left(\frac{F'\hat{F}}{T}\right).
\end{aligned}$$

with the last term on the right-hand side equal to $\frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} F\lambda_i\lambda_i'F'\hat{F}$. Denote $G =$

$\left(\frac{F'\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda_{co}'\Lambda_{co}}{N_{co}}\right)^{-1}$. After re-arranging the terms and focusing on period $t$, we have:

$$
\begin{aligned}
H^{-1}\hat{f}_t - f_t =& \frac{1}{N_{co}T}\sum_{i}^{N_{co}} G\hat{F}'X_i'(\beta-\hat{\beta})(\beta-\hat{\beta})'x_{it} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} G\hat{F}'F\lambda_i(\beta-\hat{\beta})'x_{it} \\
&+ \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} G\hat{F}'\varepsilon_i(\beta-\hat{\beta})'x_{it} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} G\hat{F}'X_i'(\beta-\hat{\beta})\lambda_i' f_t \\
&+ \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} G\hat{F}'X_i(\beta-\hat{\beta})\varepsilon_{it} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} G\hat{F}'\varepsilon_i\lambda_i' f_t \\
&+ \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} G\hat{F}'F\lambda_i\varepsilon_{it} + \frac{1}{N_{co}T}\sum_{i=1}^{N_{co}} G\hat{F}'\varepsilon_i\varepsilon_{it} \\
&= a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8,
\end{aligned}
$$

The proof of $||(\frac{F'\hat{F}}{T})^{-1}|| = O_p(1)$ is provided in Bai (2003) Proposition 1. Assumption 4 implies $||(\frac{\Lambda_{co}'\Lambda_{co}}{N_{co}})^{-1}|| = O_p(1)$, therefore, $||G|| = O_p(1)$. Also from Assumption 4, we know that $||x_{it}|| = O_p(1)$, $T^{-1/2}||X_i|| = O_p(1)$, $N_{co}^{-1/2}||\Lambda_{co}|| = O_p(1)$. Together with the facts that $T^{-1/2}||F|| = O_p(1)$ and $T^{-1/2}||\hat{F}|| = \sqrt{r}$, we have:

$$
||a_1|| \leq ||G||\frac{||\hat{F}||}{\sqrt{T}}\frac{1}{N_{co}}\sum_{i}^{N_{co}}\left(\frac{||X_i||}{\sqrt{T}}||x_{it}||\right)||\beta-\hat{\beta}||^2 = O_p(||\beta-\hat{\beta}||^2)
$$

Similarly, we can show that each of $a_2$, $a_3$, $a_4$, and $a_5$ is $O_p(\beta-\hat{\beta})$.

$$
\begin{aligned}
||a_2|| &\leq ||G||\frac{||\hat{F}||}{\sqrt{T}}\frac{||F||}{\sqrt{T}}\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}(||\lambda_i||||x_{it}||)\,||\beta-\hat{\beta}|| = O_p(||\beta-\hat{\beta}||) \\
||a_3|| &\leq ||G||\frac{||\hat{F}||}{\sqrt{T}}\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}\left(\frac{||\varepsilon_i||}{\sqrt{T}}||x_{it}||\right)||\beta-\hat{\beta}|| = O_p(||\beta-\hat{\beta}||) \\
||a_4|| &\leq ||G||\frac{||\hat{F}||}{\sqrt{T}}\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}\left(\frac{||X_i||}{\sqrt{T}}||\lambda_i' f_t||\right)||\beta-\hat{\beta}|| = O_p(||\beta-\hat{\beta}||) \\
||a_5|| &\leq ||G||\frac{||\hat{F}||}{\sqrt{T}}\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}\left(\frac{||X_i||}{\sqrt{T}}||\varepsilon_{it}||\right)||\beta-\hat{\beta}|| = O_p(||\beta-\hat{\beta}||)
\end{aligned}
$$

Moreover, $a_6$ and $a_7$ are both $O_p(N^{-1/2})$.

$$
\begin{aligned}
\|a_6\| &= \frac{1}{N_{co}T}\|G\hat{F}'\varepsilon\Lambda'_{co}f_t\| \leq \frac{1}{\sqrt{N_{co}}}\|G\|\frac{\|\hat{F}\|}{\sqrt{T}}\frac{\|\varepsilon\|}{\sqrt{T}}\frac{\|\Lambda_{co}\|}{\sqrt{N_{co}}}\|f_t\| = O_p(N_{co}^{-1/2}) \\
\|a_7\| &\leq \frac{1}{\sqrt{N_{co}}}\|G\|\frac{\|\hat{F}\|}{\sqrt{T}}\frac{\|F\|}{\sqrt{T}}\sqrt{\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}\|\lambda_i\varepsilon_{it}\|^2} = O_p(N_{co}^{-1/2})
\end{aligned}
$$

Finally, Denote $\tilde{f}_t = Gf_t,\ t = 1, 2, \cdots, T$,

$$
\begin{aligned}
a_8 &= \frac{1}{T}\sum_{s=1}^{T}\left(\tilde{f}_t\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}\varepsilon_{is}\varepsilon_{it}\right) \\
&= \frac{1}{T}\sum_{s=1}^{T}\left(\tilde{f}_t\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}[\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})]\right) - \frac{1}{T}\sum_{s=1}^{T}\left(\tilde{f}_t E(\varepsilon_{is}\varepsilon_{it})\right) \\
&= b_1 + b_2
\end{aligned}
$$

Because $E(\varepsilon_{is}\varepsilon_{it})$ is bounded according to Assumption 4.2,

$$
\|b_2\| \leq \frac{1}{\sqrt{T}}\|G\|\frac{\|F\|}{\sqrt{T}}M = O_p(T^{-1/2}).
$$

On the other hand,

$$
\|b_1\| \leq \frac{1}{\sqrt{N_{co}}}\|G\|\frac{\|\hat{F}\|}{\sqrt{T}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N_{co}}\sum_{i=1}^{N_{co}}|\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})|^2} = O_p(N_{co}^{-1/2})
$$

Therefore, $a_8 = O_p(N_{co}^{-1/2}) + O_p(T^{-1/2})$.

Because $\|\beta - \hat{\beta}\| = O_p(N_{co}^{-1}) + O_p(T^{-1}) + o_p\left((N_{co}T)^{-1/2}\right)$ according to Lemma 4,

$$
\|f_t - H^{-1}\hat{f}_t\| = O_p(\|\beta - \hat{\beta}\|) + O_p(N_{co}^{-1/2}) = O_p(N_{co}^{-1/2}) + O_p(T^{-1/2}).
$$

(ii). By subtracting $H^{-1}\hat{f}_t$ from $f_t - F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t$ and then adding it back, we have:

$$
f_t - F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t = (f_t - H^{-1}\hat{f}_t) - (F' - H^{-1}\hat{F}')\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t
$$

Because $T^{-1/2}\|F - H^{-1}\hat{F}\| = O_p(N_{co}^{-1/2}) + O_p(T^{-1/2})$ (Bai 2009, p. 1268) and $\|\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t\| =$

$O_p(T^{-1/2})$,

$$
\begin{aligned}
\|f_t - F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t\| &\leq \|f_t - H^{-1}\hat{f}_t\| + \|(F' - H^{-1}\hat{F}')\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t\| \\
&\leq \|f_t - H^{-1}\hat{f}_t\| + \|F' - H^{-1}\hat{F}'\|\|\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t\| \\
&= O_p(N_{co}^{-1/2}) + O_p(T^{-1/2})
\end{aligned}
$$

It is worth noting that if $E(\varepsilon_{is}\varepsilon_{it}) = 0$ for any $i$ and all $(s,t)$, then

$$
\|f_t - F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t\| = O_p(N_{co}^{-1/2}) + O_p(T^{-1}).
$$

**Lemma 6** $\|F'\hat{F}(\hat{F}'\hat{F})^{-1} - F^{0\prime}\hat{F}^0(\hat{F}^{0\prime}\hat{F}^0)^{-1}\| = o_p(T_0^{-1})$.

**Proof:** Denote $A = F'\hat{F}$ and $B = F'\hat{F} - F^{0\prime}\hat{F}^0$. Both are $(r \times r)$ matrices. $\|B\| = \|\sum_{s=T_0}^{T} f_s\hat{f}_s'\| = O_p(1)$. Recall $Q = \hat{F}'\hat{F} - \hat{F}^{0\prime}\hat{F}^0$ and $\hat{F}'\hat{F}/T = I_r$.

$$
\begin{aligned}
& F'\hat{F}(\hat{F}'\hat{F})^{-1} - F^{0\prime}\hat{F}^0(\hat{F}^{0\prime}\hat{F}^0)^{-1} \\
= \;& \frac{1}{T}A - (A - B)\left[\frac{1}{T}I_r + (I - \frac{1}{T}Q)^{-1}\frac{1}{T^2}Q\right] \\
= \;& \frac{1}{T}B - (A - B)(I - \frac{1}{T}Q)^{-1}\frac{1}{T^2}Q
\end{aligned}
$$

The second term on the right is $O_p(T^{-1})$ because $T^{-1}\|A - B\| = O_p(1)$. Therefore, $\|F'\hat{F}(\hat{F}'\hat{F})^{-1} - F^{0\prime}\hat{F}^0(\hat{F}^{0\prime}\hat{F}^0)^{-1}\| = O_p(T^{-1})$.

**Proposition 1 (Limit of Bias)** *Under Assumptions 1-4,* $\mathbb{E}_\varepsilon(\widehat{ATT}_t|D,X,\Lambda,F)\rightarrow ATT_t$, *in which* $ATT_t = \frac{1}{N_{tr}}\sum_{i\in\mathcal{T}}\delta_{it}$*, for all* $t > T_0$*, as both* $N_{co}$ *and* $T_0 \rightarrow \infty$.

**Proof**: Denote $i$ as the treated unit on which the treatment effect is of interest. From $Y_{it} = x_{it}'\beta + \lambda_i'f_t + \varepsilon_{it}$ and $\hat{\lambda}_i = (\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{F}^{0\prime}(Y_i^0 - X_i^0\hat{\beta})$, we have:

$$
\begin{aligned}
\hat{\delta}_{it} - \delta_{it} &= Y_{it} - \hat{Y}_{it}(0) - \delta_{it} \\
&= x_{it}'(\beta - \hat{\beta}) + (\lambda_i'f_t - \hat{\lambda}_i'\hat{f}_t) + \varepsilon_{it} \\
&= x_{it}'(\beta - \hat{\beta}) + \left\{\lambda_i'f_t - [X_i^0(\beta - \hat{\beta}) + F^0\lambda_i + \varepsilon_i^0]'\hat{F}^0(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{f}_t\right\} + \varepsilon_{it} \\
&= \left[x_{it}' - \hat{f}_t'(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{F}^{0\prime}X_i^0\right](\beta - \hat{\beta}) + \lambda_i'\left[f_t - F^{0\prime}\hat{F}^0(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{f}_t\right] + \\
&\quad \left[-\hat{f}_t'(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{F}^{0\prime}\varepsilon_i^0\right] + \varepsilon_{it} \\
&= R_{1,it} + R_{2,it} + R_{3,it} + \varepsilon_{it}, \qquad\qquad t = 1, 2, \cdots, T;\ \forall i \in \mathcal{T}.
\end{aligned}
$$

$\mathbb{E}_\varepsilon(\varepsilon_{it}|D,X,\Lambda,F) = 0$ by Assumption 2. Following a similar logic in Abadie, Diamond and Hainmueller (2010), $R_{3,it}$ can be written as:

$$R_{3,it} = -\sum_{s=1}^{T_0}\hat{f}_t'\left(\sum_{l=1}^{T_0}\hat{f}_l\hat{f}_l'\right)^{-1}\hat{f}_s\varepsilon_{is}$$

in which $\left(\sum_{l=1}^{T_0}\hat{f}_l\hat{f}_l'\right)^{-1}$ is symmetric and positive definite. Applying the Cauchy-Schwarz Inequality, we have $|\hat{f}_t'\left(\sum_{l=1}^{T_0}\hat{f}_l\hat{f}_l'\right)^{-1}\hat{f}_s| \leq O(T_0^{-1})$. Because the second moment for $\varepsilon_{it}$ exists (Assumption 3), applying the Rosenthal's Inequality, we have:

$$\mathbb{E}_\varepsilon(|R_{3,it}|^2|D,X,\Lambda,F) \leq O(T_0^{-2})\sum_{s=1}^{T_0}\mathbb{E}|\varepsilon_{is}|^2 = O(T_0^{-1}).$$

Hence, $\mathbb{E}_\varepsilon(|R_{3,it}|\ |D,X,\Lambda,F) \leq O(T_0^{-1/2})$, which means the bias from $R_{3,it}$ is bounded by a function that goes to zero as the number of pre-treatment periods grows.

Next, I investigate biases from $R_{1,it}$ and $R_{2,it}$. $R_{1,it}$ is the source of bias from imprecise estimation of $\beta$, which results in both a direct effect on the amount of bias through $x_{it}$ and an indirect effect through the estimation of the factor loading $\lambda_i$. $R_{2,it}$ is the source of bias directly from the influence of the factors $\lambda_i'f_t$. Our objective is to characterize (and bound) both $\mathbb{E}_\varepsilon(R_{1,it})$ and $\mathbb{E}_\varepsilon(R_{2,it})$. By Lemma 4 and $\|x_{it}' - \hat{f}_t'(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{F}^{0\prime}X_i^0\| = O_p(1)$ and

$$|R_{1,it}| = O_p(\|\beta - \hat{\beta}\|) = O_p(N_{co}^{-1}) + O_p(T^{-1}) + o_p\left((N_{co}T)^{-1/2}\right).$$

Using both Lemma 5 and Lemma 6, we have:

$$
\begin{aligned}
\|f_t - F^{0\prime}\hat{F}^0(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{f}_t\| &\leq \|f_t - F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t\| + \|F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t - F^{0\prime}\hat{F}^0(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{f}_t\| \\
&= O_p(N_{co}^{-1/2}) + O_p(T_0^{-1/2}).
\end{aligned}
$$

Therefore, $|R_{2,it}| \leq \|\lambda_i\|\|f_t - F^{0\prime}\hat{F}^0(\hat{F}^{0\prime}\hat{F}^0)^{-1}\hat{f}_t\| = O_p(N_{co}^{-1/2}) + O_p(T_0^{-1/2})$.

Hence, $R_{1,it} + R_{2,it}$ is bounded in probability by a function that goes to zero as $N_{co}$ and $T_0$ increases. By the moment conditions specified in Assumptions 3 and 4, $R_{1,it} + R_{2,it}$ is uniformly intergrable, therefore, convergence in probability implies convergence in means (DasGupta 2008, Ch. 6), i.e., $\mathbb{E}_\varepsilon(|R_{1,it} + R_{2,it}|\ |D, X, \Lambda, F) = O(N_{co}^{-1/2}) + O(T_0^{-1/2})$. Hence,

$$\mathbb{E}_\varepsilon(\hat{\delta}_{it} - \delta_{it}|D, X, \Lambda, F) = O(N_{co}^{-1/2}) + O(T_0^{-1/2})$$

Therefore,

$$\mathbb{E}_\varepsilon(\widehat{ATT}_t - ATT_t|D, X, \Lambda, F) = N_{tr}^{-1}\left(O(N_{co}^{-1/2}) + O(T_0^{-1/2})\right).$$

In other words, the bias of the estimator goes to zero as both $N_{co}$ and $T_0$ increase..

# Bibliography


Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *The Review of Economic Studies* pp. 1–19.

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490):493–505.

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59(2).

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak and Todd Mitton. 2013. "The Value of Connections In Turbulent Times: Evidence from the United States." Mimeo, Massachusetts Insitutte of Technology.

Angrist, Joshua D., scar Jord and Guido Kuersteiner. 2013. "Semiparametric Estimates of Monetary Policy effects: String Theory Revisited." National Bureau of Economic Research.

Bai, Jushan. 2003. "Theory for Factor Models of Large Dimensions." *Econometrica* 71(1):135–137.

Bai, Jushan. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77:1229–1279.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What to do (and not to do) with Time-Series Cross-Section Data." *American Political Science Review* 89(3):634–647.

Blackwell, Matthew and Adam Glynn. 2015. "How to Make Causal Inferences with Time-Series Cross-Sectional Data." Mimeo, Harvard Uniersity.

Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy and Steven L. Scott. 2014. "Inferring causal impact using Bayesian structural time-series models." *Annals of Applied Statistics* .

Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay. 1997. *The Econometrics of Financial Markets.* Princeton, NJ: Princeton University Press.

DasGupta, Anirban. 2008. *Asymptotic Theory of Statistics and Probability.* New York, NY: Springer.

Dube, Arindrajit and Ben Zipperer. 2015. "Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies." IZA Discussion Paper No. 8944.

Efron, Brad. 2012. "The Estimation of Prediction Error." *Journal of the American Statistical Association* 99(467):619–632.

Efron, Brad and Rob Tibshirani. 1993. *An Introduction to the Bootstrap.* New York, NY: Chapman & Halll.

Gaibulloev, Khusrav, Todd Sandler and Donggyu Sul. 2014. "Dynamic Panel Analysis under Cross-Sectional Dependence." *Political Analysis* 22(2):258–273.

Gobillon, Laurent and Thierry Magnac. 2013. "Regional Policy Evaluation: Interactive Fixed Eects and Synthetic Controls." Mimeo, Paris School of Economics.

Holland, Paul W. 1986. "945-60." *Journal of American Statistical Association* December(8):396.

Hsiao, Cheng, Steve H. Ching and Shui Ki Wan. 2012. "A Panel Data Approach for Program Evaludation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China." *Journal of Applied Econometrics* 27(5):705–740.

Imai, Kosuke and In Song Kim. 2016. "When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Panel Data.".

Kim, Dukpa and Tatsushi Oka. 2014. "Divorce Law Reforms and Divorce Rates in the USA: An Interactive Fixed-Effects Approach." *Journal of Applied Econometrics* 29(2):231–245.

Moon, Hyungsik Roger and Martin Weidner. 2013. "Dynamic Linear Panel Regression Models with Interactive Fixed Effects." Mimeo, Department of Economics, UCL.

Mora, Ricardo and Ilina Reggio. 2012. "Treatment Effect Identification Using Alternative Parallel Assumptions." Mimeo, Universidad Carlos III de Madrid.

Neyman, Jerzy. 1923. "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles." *Statistical Science* 5:465–80. Section 9 (translated in 1990).

Pang, Xun. 2010. "Modeling Heterogeneity and Serial Correlation in Binary Time-Series Cross-sectional Data: A Bayesian Multilevel Model with AR(p) Errors." *Political Analysis* (4):470–498.

Pang, Xun. 2014. "Varying Responses to Common Shocks and Complex Cross-Sectional Dependence: Dynamic Multilevel Modeling with Multifactor Error Structures for Time-Series Cross-Sectional Data." *Political Analysis* 22(4):464–496.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments n Randomized and nonrandomized Studies." *Journal of Educational Psychology* 5(66):688–701.

Stewart, Brandon. 2014. "Latent Factor Regressions for the Social Sciences." Mimeo, Harvard University.

# Chapter 2

This chapter provides empirical support for the proposed generalized synthetic control (GSC) method. I first illustrate the mechanics of the GSC estimator with a simulated sample. Then we conduct extensive Monte Carlo exercises to explore the finite sample properties of the estimator and compare it with several existing methods. Finally, I provide an empirical example in which we investigate the effect of Election Day Registration on voter turnout in the United States using the proposed method.

## A Simulated Example

I start with the following data generating process (DGP) that includes two observed time-varying covariates, two unobserved factors, and additive two-way fixed effects:

$$Y_{it} = \delta_{it}D_{it} + x_{it,1} \cdot 1 + x_{it,2} \cdot 3 + \lambda_i' f_t + \alpha_i + \xi_t + 5 + \varepsilon_{it} \tag{1}$$

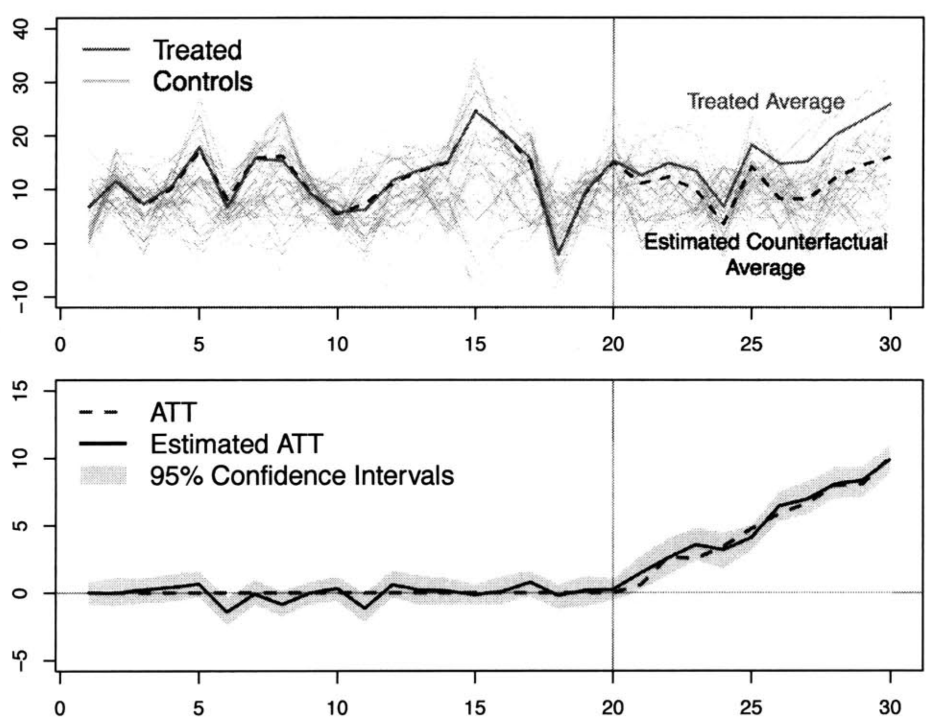
where $f_t = (f_{1t}, f_{2t})'$ and $\lambda_i = (\lambda_{i1}, \lambda_{i2})'$ are time-varying factors and unit-specific factor loadings. The covariates are (positively) correlated with both the factors and factor loadings: $x_{it,k} = 1 + \lambda_i' f_t + \lambda_{i1} + \lambda_{i2} + f_{1t} + f_{2t} + \eta_{it,k}, k = 1, 2$. The error term $\varepsilon_{it}$ and disturbances in covariates $\eta_{it,1}$ and $\eta_{it,2}$ are i.i.d. $N(0, 1)$. Factors $f_{1t}$ and $f_{2t}$, as well as time fixed effects $\xi_t$, are also i.i.d. $N(0, 1)$. The treatment and control groups consist of $N_{tr}$ and $N_{co}$ units. The treatment starts to affect the treated units at time $T_0 + 1$ and since then 10 periods are observed ($q = 10$). The treatment indicator is defined as in last chapter, i.e., $D_{it} = 1$ when $i \in \mathcal{T}$ and $t > T_0$ and $D_{it} = 0$ otherwise. The heterogeneous treatment effect is generated by $\delta_{it,t>T_0} = \bar{\delta}_t + e_{it}$, in which $e_{it}$ is i.i.d. N(0,1). $\bar{\delta}_t$ is given by: $[\bar{\delta}_{T_0+1}, \bar{\delta}_{T_0+1}, \cdots, \bar{\delta}_{T_0+10}] = [1, 2, \cdots, 10]$.

Factor loadings $\lambda_{i1}$ and $\lambda_{i2}$, as well as unit fixed effects $\alpha_i$, are drawn from uniform distributions $U[-\sqrt{3}, \sqrt{3}]$ for control units and $U[\sqrt{3} - 2w\sqrt{3}, 3\sqrt{3} - 2w\sqrt{3}]$ for treated units ($w \in [0, 1]$). This means that when $0 \leq w < 1$, (1) the random variables have variance 1; (2) the supports of factor loadings of treated and control units are not perfectly overlapped; and (3) the treatment indicator and factor loadings are positively correlated.[1]

I first illustrate the proposed method, as well as the DGP described above, with a simulated sample of $N_{tr} = 5$, $N_{co} = 45$, and $T_0 = 20$ (hence, $N = 50$, $T = 30$). $w$ is set to be 0.8, which means that the treated units are more likely to have larger factor loadings than the control units. Figure 1 visualizes the raw data and estimation results.

In the upper panel, the gray and pink lines are time series of the control and treated

Figure 1. ESTIMATED ATT FOR A SIMULATED SAMPLE
$N_{tr} = 5, N_{co} = 45, T = 30, T_0 = 10$



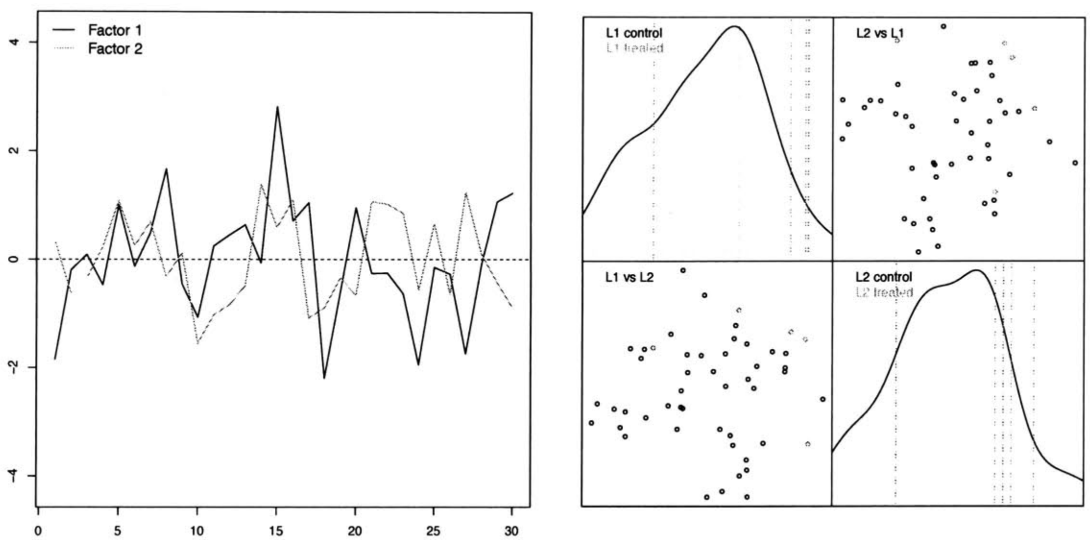

units, respectively. The bold red solid line is the average outcome of the five treated units while the bold dashed line is the average predicted outcome of the five units in the absence of the treatment. The latter is imputed using the proposed method. The lower panel of Figure 1 shows the estimated ATT (black solid line) and the true ATT (blue

[1]The DGP specified here is modified based on Bai (2009) and Gobillon and Magnac (2013).

dashed line). The 95 percent confidence intervals for the ATT are based on bootstraps of 2,000 times. It shows that the estimated average treated outcome fits the data well in pre-treatment periods and the estimated ATT is very close to the actual ATT.

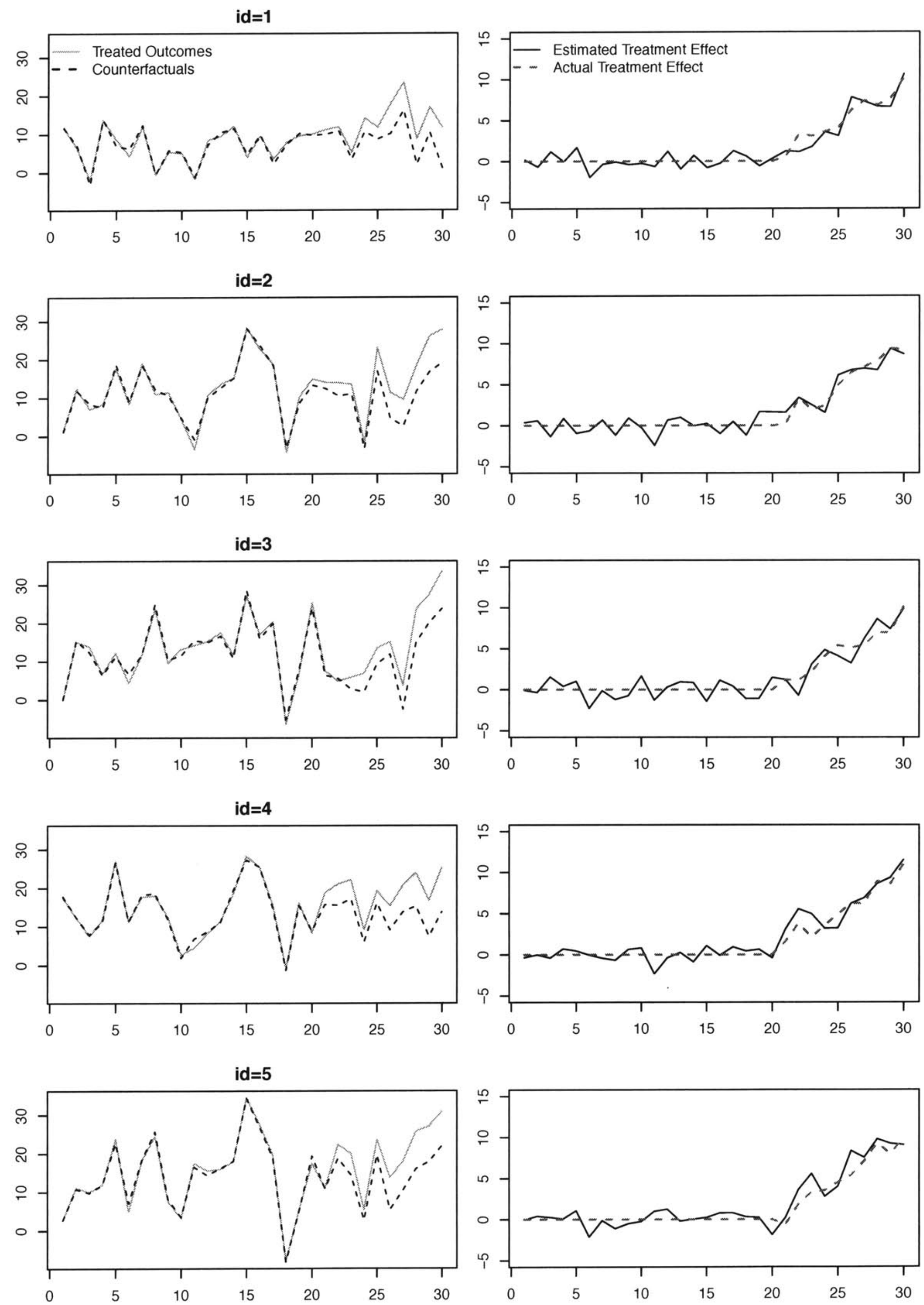
Figure 2 presents the estimated factors and factor loadings. The left panel shows the two orthogonal time-varying factors estimated by the model while the right panel shows the distributions of the estimated factor loadings for the control group (black) and the treatment group (red). The estimated counterfactual and individual treatment effect for each treat unit is shown Figure 3.

Figure 2. ESTIMATED FACTORS AND FACTOR LOADINGS: A SIMULATED SAMPLE




(a) Estimated factors (b) Estimated factor loadings

Figure 3. Estimated Individual Treatment Effect: A Simulated Sample
$N_{tr} = 5, N_{co} = 45, T = 30, T_0 = 10$



## Monte Carlo Evidence

In this section, I conduct Monte Carlo exercises to study the finite sample properties of the GSC estimator and compare it with several existing methods. including the DID estimator, the original synthetic control method, and the interactive fixed effect (IFE) estimator. I also investigate the extent to which the proposed cross-validation scheme can choose the number of factors correctly in relatively small samples, and the extent to which the GSC estimator is robust to a misspecified number of factors.

I present the Monte Carlo evidence on the finite sample properties of the GSC estimator in Table 1. As in the previous example, the treatment group is set to have five units.[2] The estimand is the ATT at time $T_0 + 5$, whose expected value equals 5. Factors are drawn at once while the factor loadings are drawn repeatedly with $w$ set to 0.5 such that treatment assignment is positively correlated with factor loadings.[3] Table 1 reports the bias, standard deviation (SD), and root mean squared error (RMSE) of $\widehat{ATT}_{T_0+5}$ from 20,000 simulations for each pair of $T_0$ and $N_{co}$.[4] It shows that the the GSC estimator has limited bias even when $T_0$ and $N_{co}$ are relatively small and the bias goes away as $T_0$ and $N_{co}$ grow. As expected, both the SD and RMSE shrink when $T_0$ and $N_{co}$ become larger..

Table 1 also reports the coverage probabilities of 95 percent confidence intervals for $\widehat{ATT}_{i,T_0+5}$ constructed by the parametric bootstrap procedure. For each pair of $T_0$ and $N_{co}$, the coverage probability is calculated based on 5,000 simulated samples, each of which is bootstrapped for 1,000 times. These numbers show that the proposed procedure can achieve the correct coverage rate even when the sample size is relatively small (e.g., $T_0 = 15$, $N_{tr} = 5$, $N_{co} = 80$).

Next, I report additional simulation results from three comparisons, which respectively

[2]Additional results with different $N_{tr}$ are shown in Table 10 in the Appendix.
[3]I fix the factors such that the biases will not be cancelled out in multiple runs.
[4]The standard deviation is defined as: $SD(\widehat{ATT}_t) = \sqrt{\mathbb{E}[\widehat{ATT}_t^{(k)} - \mathbb{E}(\widehat{ATT}_t^{(k)})]^2}$, while the root mean squared error is defined as: $RMSE(\widehat{ATT}_t) = \sqrt{\mathbb{E}(\widehat{ATT}_t^{(k)} - ATT_t^{(k)})^2}$. The superscript $(k)$ denotes the $k$-th sample.

Table 1. FINITE SAMPLE PROPERTIES AND COVERAGE RATES

| $N_{tr}$ | $N_{co}$ | $T_0$ | Bias | SD | RMSE | Coverage |
|---|---|---|---|---|---|---|
| 5 | 40 | 15 | 0.013 | 0.847 | 0.722 | 0.946 |
| 5 | 80 | 15 | 0.007 | 0.769 | 0.624 | 0.950 |
| 5 | 120 | 15 | 0.004 | 0.741 | 0.590 | 0.949 |
| 5 | 200 | 15 | 0.001 | 0.715 | 0.556 | 0.949 |
| 5 | 40 | 30 | -0.012 | 0.813 | 0.674 | 0.945 |
| 5 | 80 | 30 | -0.006 | 0.734 | 0.580 | 0.948 |
| 5 | 120 | 30 | -0.004 | 0.705 | 0.546 | 0.949 |
| 5 | 200 | 30 | -0.007 | 0.681 | 0.521 | 0.948 |
| 5 | 40 | 50 | -0.015 | 0.782 | 0.643 | 0.947 |
| 5 | 80 | 50 | -0.002 | 0.715 | 0.558 | 0.948 |
| 5 | 120 | 50 | -0.006 | 0.695 | 0.525 | 0.949 |
| 5 | 200 | 50 | 0.003 | 0.679 | 0.502 | 0.949 |

show that (1) the GSC estimator has much less bias than the DID estimator in the presence of unobserved, decomposable time-varying confounders; (2) it is more efficient than the original synthetic control estimator; and (3) it has significantly less bias than the IFE estimator when the treatment effect is heterogeneous. These results are under the premise of correct model specifications. To address the concern that the GSC method is sensitive to specifications, I conduct two additional sets of simulations. First, I show that the cross-validation scheme described in the last chapter is able to choose the number of factors correctly most of the time when the sample is large enough. Then I show that the GSC estimates remain unbiased even when number of factors included in the model exceeds the correct number.

**Unobserved confounders and a comparison with the DID estimator.** Table 2 compares the GSC and DID estimates. It shows that when the treatment is randomly assigned ($w = 1$), both methods have limited bias while the GSC method is more efficient than DID. However, when the treatment is not randomly assigned ($w < 1$), the DID estimator yields huge bias while the bias of the GSC estimates remains small. These results suggest that, given sufficient data, the GSC estimator is both more robust and often more efficient than the commonly used DID estimator in estimating the treatment effect when decomposable, time-varying confounders exist.

Table 2. COMPARISION WITH THE DIFFERENCE-IN-DIFFERENCES ESITMATOR

| | | | | GSC | | | DID | | |
|---|---|---|---|---|---|---|---|---|---|
| $T_0$ | $N_{tr}$ | $N_{co}$ | $w$ | Bias | SD | RMSE | Bias | SD | RMSE |
| 15 | 10 | 40 | 1.00 | 0.010 | 0.526 | 0.418 | 0.007 | 0.629 | 0.542 |
| 15 | 20 | 40 | 1.00 | 0.004 | 0.386 | 0.318 | 0.006 | 0.466 | 0.410 |
| 15 | 30 | 40 | 1.00 | 0.000 | 0.332 | 0.283 | -0.005 | 0.400 | 0.361 |
| 15 | 40 | 40 | 1.00 | 0.002 | 0.310 | 0.263 | 0.001 | 0.376 | 0.337 |
| 15 | 10 | 40 | 0.75 | 0.004 | 0.577 | 0.480 | -0.357 | 0.623 | 0.641 |
| 15 | 20 | 40 | 0.75 | 0.004 | 0.455 | 0.399 | -0.349 | 0.454 | 0.532 |
| 15 | 30 | 40 | 0.75 | -0.002 | 0.412 | 0.369 | -0.355 | 0.398 | 0.501 |
| 15 | 40 | 40 | 0.75 | -0.002 | 0.386 | 0.349 | -0.342 | 0.370 | 0.476 |
| 15 | 10 | 40 | 0.50 | 0.012 | 0.701 | 0.625 | -0.629 | 0.610 | 0.815 |
| 15 | 20 | 40 | 0.50 | 0.013 | 0.603 | 0.560 | -0.597 | 0.450 | 0.718 |
| 15 | 30 | 40 | 0.50 | -0.008 | 0.567 | 0.534 | -0.598 | 0.398 | 0.694 |
| 15 | 40 | 40 | 0.50 | 0.001 | 0.562 | 0.543 | -0.595 | 0.361 | 0.679 |

**Note:** This table compares performances of the generalized synthetic control estimator (GSC) and the difference-in-differences estimator (DID) for $ATT_{T_0+5}$ with different combinations of $N_{tr}$ and $w$, a parameter that characterizes the overlap of support of factor loadings (including unit fixed effects) between the treated and control units. Each set of numbers is based on 20,000 simulated samples. The data generating process (DGP) is as follows:

$$Y_{it} = \delta_{it}D_{it} + x_{it,1} \cdot 1 + x_{it,2} \cdot 3 + \lambda_{i1}f_{1t} + \lambda_{i2}f_{2t} + \alpha_i + \xi_t + 5 + \varepsilon_{it}.$$

(1) The treatment indicator $D_{it}$ equal 1 when $i \in \mathcal{T}$ and $t > T_0$ and 0 otherwise. Heterogenous treatment effects are generated by $\delta_{it,t>T_0} = \bar{\delta}_t + e_{it}$, in which $e_{it}$ is i.i.d. N(0,1) and $[\bar{\delta}_{T_0+1}, \bar{\delta}_{T_0+1}, \cdots, \bar{\delta}_{T_0+10}] = [1, 2, \cdots, 10]$. Hence, $\mathbb{E}[ATT_{T_0+5}] = 5$. (2) The regressors are generated by: $x_{it,k} = 1 + \sum_{s=1}^{2}(\frac{1}{2}\lambda_{is}f_{st} + \frac{1}{4}\lambda_{is} + \frac{1}{4}f_{st}) + \eta_{it,k}$, $k = 1, 2$ in which $\eta_{it,k}$ is i.i.d. $N(0,1)$. (3) The factors $f_{st}, s = 1, 2$ and time fixed effects $\xi_t$, are i.i.d. $N(0,1)$ for all $t$ and are drawn only once. (4) The factor loadings $\lambda_{is}, s = 1, 2, \cdots, r$ and unit fixed effects $\alpha_i$ are i.i.d. $U[-\sqrt{3}, \sqrt{3}]$ for $i \in \mathcal{C}$ and i.i.d. $U[\sqrt{3} - 2\sqrt{3}w, 3\sqrt{3} - 2\sqrt{3}w]$ for $i \in \mathcal{T}$, which means treated and control units have common support when $w = 1$ any lack almost any common support when $w = 0$. (5) The error term $\varepsilon_{it}$ is i.i.d. $N(0,1)$.

**Common support and a comparison with the synthetic control method.** The synthetic control method proposed by Abadie, Diamond and Hainmueller (2010) requires that both covariates and factor loadings of the treated unit is in the convex hull of those of the donors from the control group. The method may fail to construct a synthetic control unit when this requirement is not met. In this way, it safeguards against unwarranted extrapolations that may lead to biased estimates of the treatment effect.

The GSC method, however, does not have this requirement–in this sense, it is less conservative in terms of imputing treated counterfactuals. First, like DID, it allows for an intercept shift when additive unit fixed effects are assumed. Second, it incorporates observable covariates by imposing parametric assumptions. Third, in the lack of common support of factor loadings between the treated and control groups, it extrapolates the influence of the factors on the treated outcome based on the assumed model. When the model is correct, the GSC estimator is expected to be more efficient than the original synthetic control method because it potentially uses more information: (1) no control units are discarded and even negative correlations between the treated and control units are used for the prediction of treated counterfactuals; (2) when the model specifies more than one unobserved factors, a control unit at different time periods is assigned different weights. To be more precise, the control units are first decomposed into several components (factors) and these components are re-weighted to produce treated counterfactuals. When the model is incorrect, however, such extrapolations cause biases.[5] Therefore, when applying the GSC method, it is helpful to plot the estimated factors and factor loadings to avoid excessive extrapolations.

Table 3 compares the GSC method with the original synthetic control method (labelled as **Synth**) and confirms our expectation. It shows that when the model is correct and the treated and control units shares a common support of factor loadings, both methods have limited bias while the GSC estimator is more efficient (there is a 0.8-1.6 percent chance that the original synthetic control method may fail to construct a synthetic control

[5]The GSC method safeguards against this risk to some extent by incorporating specification error into the estimation of uncertainties: the prediction error becomes larger when the model is mis-specified. This is because when the model is mis-specified, it is likely to do a poor job in predicting outcomes of the control units in the cross-validation procedure.

unit, though). As the overlap of support between treated and control units diminishes, significant bias shows up for the original synthetic control method while the bias of GSC method remains small.

**Heterogeneous treatment effect and a comparison with the IFE estimator**. The GSC estimator is a bias-correction method for IFE models when the treatment effect is heterogeneous across units. When the treatment effect is constant, the IFE estimator is more efficient because it uses information of both the treatment and control groups to estimate covariate coefficients and factors while the GSC method uses the control group information only. When the treatment effect is heterogeneous, however, using IFE models that assume constant treatment effect will lead to biased estimates because heterogeneities of the treatment effect will cause inconsistent estimation of the factor space.

Results from Monte Carlo exercises are consistent with the above intuition. Table 4 compares the performances of the GSC estimator and the IFE estimator. The DGP is as specified in Equation (1) with $w$ set to 0.5 as usual (factor loadings are positively correlated with treatment assignment). The IFE model to be estimated uses the following specification:

$$Y_{it} = \sum_{t=T_0+1}^{T} \delta_t D_{it} + x'_{it}\beta + \lambda'_i F_t + \alpha_i + \xi_t + \varepsilon_{it},$$

which allows the treatment effects to be different over time. Table 4 shows that (1) when the treatment effect is constant across units, both estimators have limited bias and the IFE estimator out-performs the GSC estimator in terms of efficiency as the treatment group becomes larger; (2) when the treatment effect is heterogeneous, the bias of the GSC estimates remains small while the bias of of the IFE estimates increases as the variance of the treatment effect grows.

Table 3. COMPARISON WITH THE SYNTHETIC CONTROL ESTIMATOR (ADH 2010)

| | | | | | GSC | | | | Synth | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T_0$ | $N_{tr}$ | $N_{co}$ | $r$ | $w$ | Bias | SD | RMSE | Fail | Bias | SD* | RMSE* | Fail |
| 15 | 1 | 40 | 1 | 1.00 | -0.010 | 1.494 | 1.107 | 0.000 | -0.011 | 1.739 | 1.417 | 0.013 |
| 15 | 1 | 40 | 2 | 1.00 | 0.000 | 1.571 | 1.190 | 0.000 | -0.022 | 2.029 | 1.738 | 0.008 |
| 15 | 1 | 40 | 3 | 1.00 | -0.003 | 1.581 | 1.220 | 0.000 | 0.021 | 2.368 | 2.151 | 0.014 |
| 15 | 1 | 40 | 4 | 1.00 | 0.013 | 1.610 | 1.253 | 0.000 | -0.013 | 2.345 | 2.122 | 0.016 |
| 15 | 1 | 40 | 2 | 0.75 | 0.014 | 1.595 | 1.234 | 0.000 | 0.707 | 2.096 | 1.975 | 0.016 |
| 15 | 1 | 40 | 2 | 0.50 | 0.026 | 1.602 | 1.257 | 0.000 | 1.331 | 2.327 | 2.489 | 0.014 |
| 15 | 1 | 40 | 2 | 0.25 | 0.000 | 1.729 | 1.391 | 0.000 | 1.630 | 2.492 | 2.803 | 0.016 |
| 15 | 1 | 40 | 2 | 0.00 | 0.033 | 1.822 | 1.521 | 0.000 | 2.127 | 2.610 | 3.199 | 0.012 |

**Note:** This table compares performances of the generalized synthetic control estimator (GSC)and the original synthetic control estimator (`Synth`) for $ATT_{T_0+5}$ with different combinations of $r$, the number of factors, and $w$, a parameter that characterizes the overlap of support of factor loadings (including unit fixed effects) between the treated and control units. Each set of numbers is based on 20,000 simulated samples. The data generating process (DGP) is as follows:

$$Y_{it} = \delta_{it} D_{it} + x_{it,1} \cdot 1 + x_{it,2} \cdot 3 + \sum_{s=1}^{r} \lambda_{is} f_{st} + \alpha_i + \xi_t + 5 + \varepsilon_{it}.$$

(1) The treatment indicator $D_{it}$ equal 1 when $i \in \mathcal{T}$ and $t > T_0$ and 0 otherwise. Heterogenous treatment effects are generated by $\delta_{it,t>T_0} = \bar{\delta}_t + e_{it}$, in which $e_{it}$ is i.i.d. N(0,1) and $[\bar{\delta}_{T_0+1}, \bar{\delta}_{T_0+1}, \cdots, \bar{\delta}_{T_0+10}] = [1, 2, \cdots, 10]$. Hence, $\mathbb{E}[ATT_{T_0+5}] = 5$. (2) The regressors are generated by: $x_{it,k} = 1 + \frac{1}{2}\sum_{s=1}^{2} \lambda_{is} f_{st} + \frac{1}{4}\lambda_{i1} + \frac{1}{4}\lambda_{i2} + \frac{1}{4} f_{1t} + \frac{1}{4} f_{2t} + \eta_{it,k}$, $k = 1, 2$ in which $\eta_{it,k}$ is i.i.d. $N(0,1)$. (3) The factors $f_{st}, s = 1, 2, \cdots, r$ and time fixed effects $\xi_t$, are i.i.d. $N(0,1)$ for all $t$ and are drawn only once. (4) The factor loadings $\lambda_{is}, s = 1, 2, \cdots, r$ and unit fixed effects $\alpha_i$ are i.i.d. $U[-\sqrt{3}, \sqrt{3}]$ for $i \in \mathcal{C}$ and i.i.d. $U[\sqrt{3} - 2\sqrt{3}w, 3\sqrt{3} - 2\sqrt{3}w]$ for $i \in \mathcal{T}$, which means treated and control units have common support when $w = 1$ any lack almost any common support when $w = 0$. (5) The error term $\varepsilon_{it}$ is i.i.d. $N(0,1)$.

Table 4. COMPARISION WITH THE INTERACTIVE FIXED-EFFECT ESTIMATOR

| | | | | GSC | | | IFE | | |
|---|---|---|---|---|---|---|---|---|---|
| $T_0$ | $N_{co}$ | $N_{tr}$ | $var(\delta_{it})$ | Bias | SD | RMSE | Bias | SD | RMSE |
| 15 | 40 | 1 | 0 | 0.012 | 1.280 | 1.280 | 0.008 | 1.297 | 1.297 |
| 15 | 40 | 5 | 0 | 0.012 | 0.725 | 0.725 | 0.009 | 0.723 | 0.723 |
| 15 | 40 | 20 | 0 | -0.000 | 0.566 | 0.566 | -0.005 | 0.501 | 0.501 |
| 15 | 40 | 40 | 0 | 0.008 | 0.530 | 0.530 | 0.000 | 0.422 | 0.422 |
| 15 | 40 | 1 | 25 | -0.000 | 5.155 | 1.260 | -0.003 | 5.159 | 1.275 |
| 15 | 40 | 5 | 25 | 0.003 | 2.358 | 0.721 | -0.009 | 5.654 | 5.187 |
| 15 | 40 | 20 | 25 | 0.011 | 1.262 | 0.564 | 0.923 | 8.344 | 8.311 |
| 15 | 40 | 40 | 25 | 0.007 | 0.950 | 0.531 | 1.618 | 9.426 | 9.530 |
| 15 | 40 | 10 | 4 | 0.006 | 0.885 | 0.615 | 0.006 | 0.962 | 0.725 |
| 15 | 40 | 10 | 16 | 0.006 | 1.403 | 0.616 | -0.035 | 3.958 | 3.749 |
| 15 | 40 | 10 | 36 | 0.006 | 2.000 | 0.617 | 0.862 | 9.277 | 9.137 |
| 15 | 40 | 10 | 64 | 0.007 | 2.606 | 0.620 | 1.384 | 12.269 | 12.101 |

**Note:** This table compares performances of the generalized synthetic control estimator (GSC) and the interactive fixed-effect estimator (IFE) for $ATT_{T_0+5}$ with different combinations of $N_{tr}$ and $var(\delta_{it})$, the variance of individual treatment effects. Each set of numbers is based on 20,000 simulated samples. The data generating process (DGP) is as follows:

$$Y_{it} = \delta_{it}D_{it} + x_{it,1} \cdot 1 + x_{it,2} \cdot 3 + \lambda_{i1}f_{1t} + \lambda_{i2}f_{2t} + \alpha_i + \xi_t + 5 + \varepsilon_{it}.$$

(1) The treatment indicator $D_{it}$ equal 1 when $i \in \mathcal{T}$ and $t > T_0$ and 0 otherwise. Heterogenous treatment effects are generated by $\delta_{it,t>T_0} = \bar{\delta}_t + e_{it}$, in which $e_{it}$ is i.i.d. $N(0, \sigma^2)$ and $[\bar{\delta}_{T_0+1}, \bar{\delta}_{T_0+1}, \cdots, \bar{\delta}_{T_0+10}] = [1, 2, \cdots, 10]$. Hence, $var(\delta_{it}) = \sigma^2$ and $\mathbb{E}[ATT_{T_0+5}] = 5$. (2) The regressors are generated by: $x_{it,k} = 1 + \frac{1}{2}\sum_{s=1}^{2} \lambda_{is}f_{st} + \frac{1}{4}\lambda_{i1} + \frac{1}{4}\lambda_{i2} + \frac{1}{4}f_{1t} + \frac{1}{4}f_{2t} + \eta_{it,k}$, $k = 1, 2$ in which $\eta_{it,k}$ is i.i.d. $N(0, 1)$. (3) The factors $f_{st}$, $s = 1, 2$ and time fixed effects $\xi_t$, are i.i.d. $N(0, 1)$ for all $t$ and are drawn only once. (4) The factor loadings $\lambda_{is}, s = 1, 2$ and unit fixed effects $\alpha_i$ are i.i.d. $U[-\sqrt{3}, \sqrt{3}]$ for $i \in \mathcal{C}$ and i.i.d. $U[0, 2\sqrt{3}]$ for $i \in \mathcal{T}$, which produces positive correlations between the treatment indicator, the factor loadings, and the regressors. (5) The error term $\varepsilon_{it}$ is i.i.d. $N(0, 1)$.

**Choosing the number of factors and robustness to model specifications.** So far I have shown that when the factor model is correctly specified, the GSC estimator performs well in small samples and have advantages over the DID estimator, the original synthetic control method, and the IFE estimator under various circumstances. Finally, I investigate whether the cross-validation scheme proposed earlier in this paper is able to select the correct number of factors when it is unknown and whether the GSC method is robust to a misspecified number of factors.

First, I conduct simulations using the same DGP specified in Equation (1) (with $w = 0.5$) and let the algorithm choose the number of factors automatically. Table 5 shows the percentage of correct choices of the number of factors with different sample size from 5,000 simulations for each case. It suggests that when the sample is reasonably large, with a high chance the cross-validation scheme can choose the number of factors correctly. For example, when $T_0 = 30$, $N_{co} = 40$ and $N_{tr} = 5$, the cross-validation algorithm correctly chooses the number of factors 92.1% of the time; the number increases to 98.5% when $N_{tr} = 20$. Note that the number of treated units $N_{tr}$ matters because a larger treatment group provide more data for validation. In Figure 4, I plot the MSPE of all 5,000 simulations and their median MSPE in each case for four combinations of $T_0$, $N_{co}$, and $N_{tr}$. The median MSPE is always the lowest when the number of factors is correct, i.e., $r = 2$.

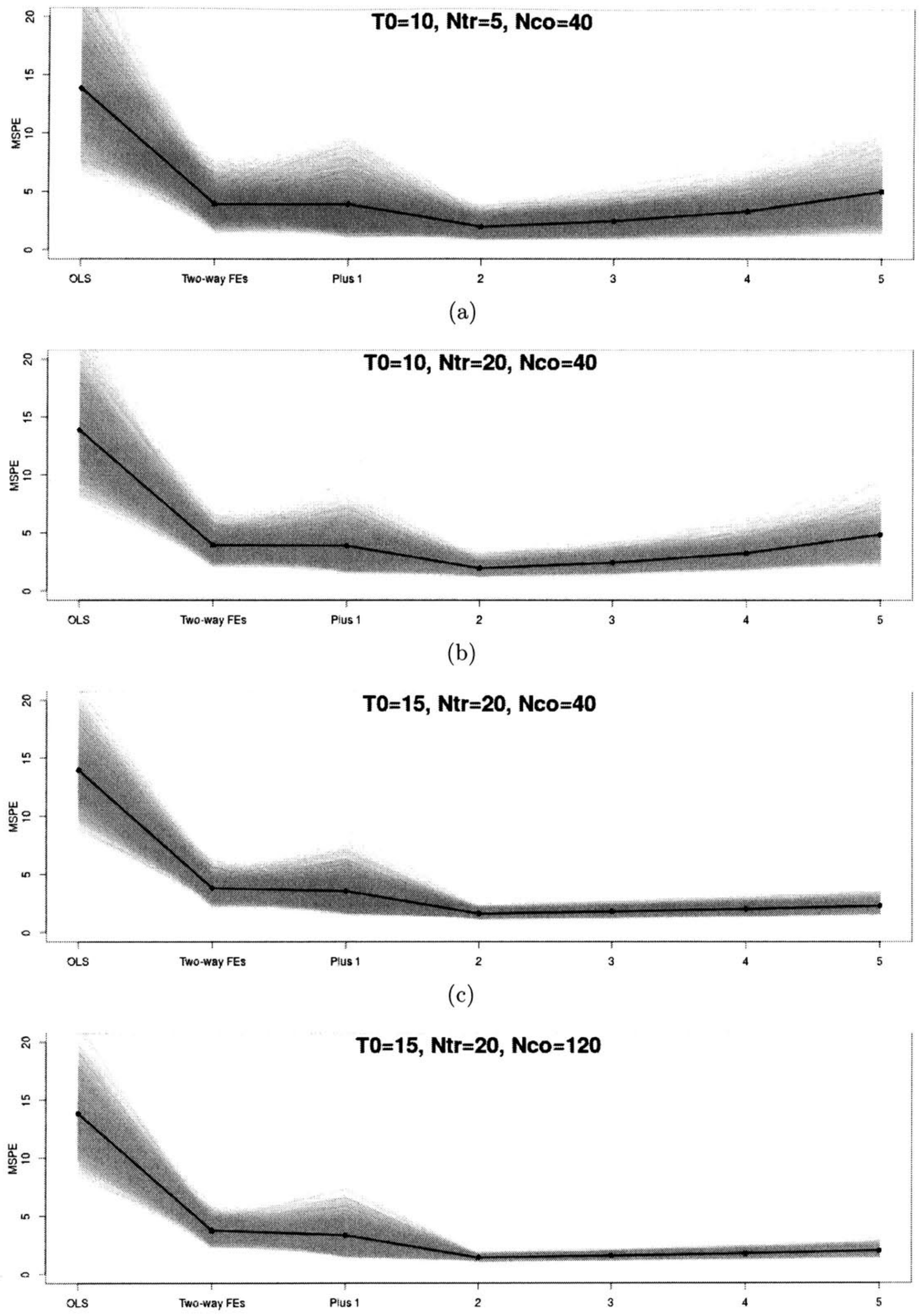
Table 5. CHOICES OF THE NUMBER OF FACTORS

| | | $N_{tr} = 5$ | $N_{tr} = 20$ | $N_{tr} = 40$ |
|---|---|---|---|---|
| $T_0$ | $N_{co}$ | $r\checkmark$ | $r\checkmark$ | $r\checkmark$ |
| 10 | 40 | 0.801 | 0.938 | 0.953 |
| 30 | 40 | 0.921 | 0.985 | 0.991 |
| 50 | 40 | 0.943 | 0.990 | 0.996 |
| 15 | 40 | 0.879 | 0.976 | 0.991 |
| 15 | 80 | 0.896 | 0.992 | 0.998 |
| 15 | 120 | 0.895 | 0.995 | 0.999 |

**Note:** Each number is based on 5,000 simulated samples.

Second, I test whether the method is robust to a mis-specified number of factors. Table 6 shows that, if the number of factors included in the model is larger than the correct number (equal to 2 as before), the bias of the estimated treatment effect will shrink to zero as the sample size becomes large; as expected, the treatment effect is less precisely estimated than with the correct number of factors. However, if the number of factors is less than the correct number, significant bias of the treatment effect remains even with large samples.

Figure 4. Choice of the Number of Factors: Four Cases




**Note:** This set of figures shows the MSPEs of six models, including pooled OLS, the two-way fixed effects model, and the GSC method with 1 to 5 factors (shown on the x-axis), under four different scenarios. 5,000 simulations are conducted for each scenario. Results from all simulations are represented with gray lines. The black solid line shows the median MSPE of 5,000 simulations with each model.

Table 6. Robustness to Misspecifications

| $T_0$ | $N_{co}$ | $r=0$ Bias | SD | RMSE | $r=1$ Bias | SD | RMSE | $r=2$ (correct) Bias | SD | RMSE | $r=3$ Bias | SD | RMSE | $r=4$ Bias | SD | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 40 | -0.724 | 0.844 | 1.017 | 0.819 | 0.931 | 1.157 | 0.008 | 0.849 | 0.719 | 0.008 | 0.878 | 0.755 | 0.009 | 0.915 | 0.796 |
| 15 | 80 | -0.737 | 0.825 | 1.011 | 0.972 | 0.823 | 1.190 | -0.007 | 0.769 | 0.623 | -0.007 | 0.795 | 0.654 | -0.006 | 0.823 | 0.688 |
| 15 | 120 | -0.733 | 0.816 | 1.002 | 1.041 | 0.787 | 1.229 | 0.009 | 0.738 | 0.590 | 0.008 | 0.761 | 0.619 | 0.007 | 0.786 | 0.650 |
| 15 | 200 | -0.735 | 0.813 | 0.999 | 1.075 | 0.771 | 1.245 | 0.001 | 0.714 | 0.557 | -0.000 | 0.734 | 0.582 | -0.001 | 0.759 | 0.613 |
| 30 | 40 | -0.573 | 0.828 | 0.906 | 0.551 | 1.063 | 1.116 | -0.014 | 0.806 | 0.672 | -0.015 | 0.815 | 0.684 | -0.016 | 0.826 | 0.697 |
| 30 | 80 | -0.562 | 0.820 | 0.885 | 0.806 | 0.929 | 1.146 | -0.006 | 0.734 | 0.579 | -0.005 | 0.745 | 0.593 | -0.006 | 0.754 | 0.604 |
| 30 | 120 | -0.575 | 0.815 | 0.889 | 0.901 | 0.863 | 1.166 | -0.006 | 0.710 | 0.548 | -0.008 | 0.719 | 0.560 | -0.008 | 0.729 | 0.573 |
| 30 | 200 | -0.566 | 0.807 | 0.878 | 1.001 | 0.820 | 1.213 | 0.009 | 0.693 | 0.528 | 0.008 | 0.701 | 0.538 | 0.008 | 0.711 | 0.550 |
| 50 | 40 | -0.277 | 0.802 | 0.723 | -0.379 | 1.057 | 1.035 | -0.020 | 0.787 | 0.646 | -0.019 | 0.793 | 0.653 | -0.020 | 0.800 | 0.660 |
| 50 | 80 | -0.273 | 0.788 | 0.703 | -0.439 | 1.041 | 1.036 | -0.009 | 0.720 | 0.558 | -0.009 | 0.725 | 0.565 | -0.010 | 0.731 | 0.572 |
| 50 | 120 | -0.276 | 0.776 | 0.691 | -0.479 | 1.008 | 1.024 | -0.003 | 0.688 | 0.521 | -0.003 | 0.691 | 0.526 | -0.003 | 0.696 | 0.532 |
| 50 | 200 | -0.280 | 0.777 | 0.698 | -0.563 | 0.982 | 1.043 | -0.004 | 0.673 | 0.502 | -0.004 | 0.678 | 0.508 | -0.003 | 0.683 | 0.514 |

**Note:** This table shows the bias, standard deviation, and root mean square error of $\widehat{ATT}_{T_0+5}$ with different combinations of $N_{co}$ and $T_0$ ($N_{tr}$ is set to be 5) when the number of factors included in the model is set differently. Each set of numbers is based on 20,000 simulated samples. The data generating process (DGP) is as follows:

$$Y_{it} = \delta_{it}D_{it} + x_{it,1}\cdot 1 + x_{it,2}\cdot 3 + \lambda_{i1}f_{1t} + \lambda_{i2}f_{2t} + \alpha_i + \xi_t + 5 + \varepsilon_{it}.$$

(1) The treatment indicator $D_{it}$ equal 1 when $i \in \mathcal{T}$ and $t > T_0$ and 0 otherwise. Heterogenous treatment effects are generated by $\delta_{it,t>T_0} = \bar{\delta}_t + e_{it}$, in which $e_{it}$ is i.i.d. N(0,1) and $[\bar{\delta}_{T_0+1}, \bar{\delta}_{T_0+1}, \cdots, \bar{\delta}_{T_0+10}] = [1, 2, \cdots, 10]$. Hence, $\mathbb{E}[ATT_{T_0+5}] = 5$. (2) The regressors are generated by: $x_{it,k} = 1+\frac{1}{2}\sum_{s=1}^{2}\lambda_{is}f_{st}+\frac{1}{4}\lambda_{i1}+\frac{1}{4}\lambda_{i2}+\frac{1}{4}f_{1t}+\frac{1}{4}f_{2t}+\eta_{it,k}$, $k = 1, 2$ in which $\eta_{it,k}$ is i.i.d. $N(0,1)$. (3) The factors $f_{st}, s = 1, 2$ and time fixed effects $\xi_t$, are i.i.d. $N(0,1)$ for all $t$ and are drawn only once. (4) The factor loadings $\lambda_{is}, s = 1, 2$ and unit fixed effects $\alpha_i$ are i.i.d. $U[-\sqrt{3}, \sqrt{3}]$ for $i \in \mathcal{C}$ and i.i.d. $U[0, 2\sqrt{3}]$ for $i \in \mathcal{T}$, which produces positive correlations between the treatment indicator, the factor loadings, and the regressors. (5) The error term $\varepsilon_{it}$ is i.i.d. $N(0,1)$.

## Election Day Registration on Voter Turnout

In this section, I illustrate the GSC method with an empirical example that investigates the effect of Election Day Registration (EDR) laws on voter turnout in the United States. Voting in the United States usually takes two steps. Except in North Dakota, where no registration is needed, eligible voters throughout the country must register prior to casting their ballots. Registration, which often requires a separate trip from voting, is widely regarded as a substantial cost of voting and a culprit of low turnout rates before the 1993 National Voter Registration Act (NVRA) was enacted (e.g. Highton 2004). Against this backdrop, EDR is a reform that allows eligible voters to register on Election Day when they arrive at polling stations. In the mid-1970s, Maine, Minnesota, and Wisconsin were the first adopters of this reform in the hopes of increasing voter turnout; while Idaho, New Hampshire, and Wyoming established EDR in the 1990s as a strategy to opt out the NVRA (Hanmer 2009). Before the 2012 presidential election, three other states,

Table 7. State EDR Laws

| State | Enacted | Took effect |
|---|---|---|
| Maine | 1973 | 1976 |
| Minnesota | 1974 | 1976 |
| Wisconsin | 1975 | 1976 |
| Wyoming | 1994 | 1996 |
| Idaho | 1994 | 1996 |
| New Hampshire | 1996 | 1996 |
| Montana | 2005 | 2008 |
| Iowa | 2007 | 2008 |
| Connecticut | 2012 | 2012 |

Montana, Iowa, and Connecticut, passed laws to enact EDR, adding the number of states having EDR laws to nine. Table 7 lists the years during which EDR laws were enacted and first took effect in presidential elections for the 9 treated states.

Most existing studies based on individual-level cross-sectional data, such as the Current Population Surveys and the National Election Surveys, suggest that EDR laws increase turnout (the estimated effect varies from 5 to 14 percentage points).[6] These studies do not provide compelling evidence of a causal effect of EDR laws because the research designs they use are insufficient to address the problem that states self-select their systems of registration laws. "Registration requirements did not descend from the skies," as Dean Burnham puts it (1980, p. 69). A few studies employ time-series or TSCS analysis to address the identification problem.[7] However, Keele and Minozzi (2013) cast doubts on these studies and suggest that the "parallel trends" assumption may not hold, as we will also demonstrate below.

In the following analysis, I use state-level voter turnout data for presidential elections from 1920 to 2012.[8] The turnout rates are calculated with total ballots counted in a presidential election in a state as the numerator and the state's voting-age population (VAP) as the denominator.[9] Alaska and Hawaii are not included in the sample since they were not states until 1959. North Dakota is also dropped since no registration is required. As mentioned above, up to the 2012 presidential election, 9 states had adopted EDR laws (hereafter referred to as *treated*) and the rest 38 states had not (referred to as *controls*). Figure 5 shows the raw data of state-level turnout rates (%) in US presidential elections in 47 states from 1920 to 2012. Turnout rates of 38 states that had not adopted EDR laws (controls) are in gray. For the 9 states in which EDR laws took effect before 2012 (treated), the pre- and post-EDR periods are represented by blue solid lines and red dashed lines, respectively. As is shown in the figure and has been pointed out by

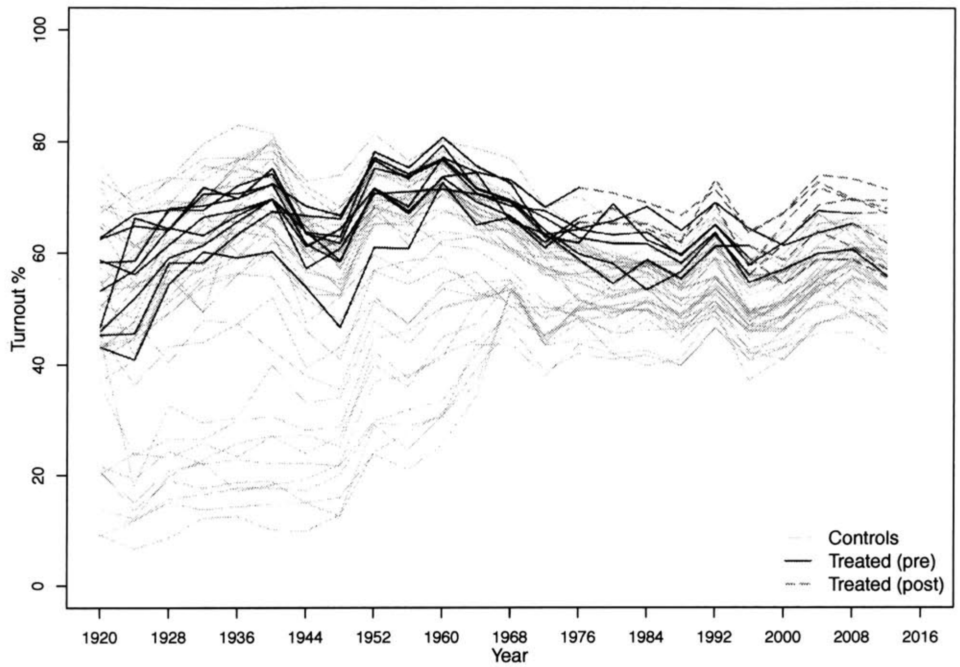
[6]See Wolfinger and Rosenstone (1980), Mitchell and Wlezien (1995), Rhine (1992), Highton (1997), Timpone (1998, 2002), Huang and Shields (2000), Brians and Grofman (2001), Hanmer (2009), Burden et al. (2009), Cain, Donovan and Tolbert (2011), Teixeira (2011) for examples. The results are especially consistent for the three early adopters, Maine, Minnesota, and Wisconsin.

[7]See, for example, Fenster (1994), King and Wambeam (1995), Knack and White (2000), Knack (2001), Neiheisel and Burden (2012), Springer (2014).

[8]The data from 1920 to 2000 are from Springer (2014). The data from 2004 to 2012 are from The United States Election Project, `http://www.electproject.org/`. Indicators of other registration laws, including universal mail-in registration and motor voter registration, also come from Springer (2014), with a few supplements.

[9]I do not use the voting-eligible population (VEP) as the denominator because they are not available in early years.

Figure 5. EDR AND STATE-LEVEL TURNOUT: RAW DATA




many, turnout rates are in general higher in states that have EDR laws than states that have not, but this does not necessarily imply a causal relationship between EDR laws and voter turnout.

First, I conduct a DID analysis using a standard two-way fixed effects model. The results are shown in Table 8 columns (1) and (2). Standard errors are produced by non-parametric bootstraps (blocked at the state level) of 1,000 times. In column (1), only the EDR indicator is included, while in column (2), I additionally control for indicators of universal mail-in registration and motor voter registration. The estimated coefficients of EDR laws are 0.87 and 0.78 percent using the two specifications, respectively, with standard errors around 3 percent.

The DID model presented in Table 8 assumes a constant treatment effect both across states and over time. I relax this assumption by allowing the treatment effect to be different across time (but the effects are assumed to be the same across states when the number of terms before or after EDR laws took effect is the same). In other words, I estimate the dynamic effect of EDR laws on voter turnout by interacting the indicator of

Table 8. THE EFFECT OF EDR ON VOTER TURNOUT

| *Outcome variable* | Voter Turnout % | | | |
|---|---|---|---|---|
| | DID | | GSC | |
| | (1) | (2) | (3) | (4) |
| Election Day Registration | 0.87 | 0.78 | 5.13 | 4.90 |
| | (3.01) | (3.44) | (2.20) | (2.26) |
| Universal Mail-in Registration | | -0.94 | | 0.15 |
| | | (1.78) | | (0.80) |
| Motor Voter Registration | | -0.20 | | -1.05 |
| | | (1.51) | | (0.79) |
| State fixed effects | x | x | x | x |
| Year fixed effects | x | x | x | x |
| Unobserved factors | N/A | N/A | 2 | 2 |
| Observations | 1,128 | 1,128 | 1,128 | 1,128 |
| Treated states | 9 | 9 | 9 | 9 |
| Control states | 38 | 38 | 38 | 38 |

**Note:** Standard errors in columns (1) and (2) are based on non-parametric boostraps (blocked at the state level) of 2,000 times. Standard errors in columns (3) and (4) are based on parametric bootstraps (blocked at the state level) of 2,000 times.

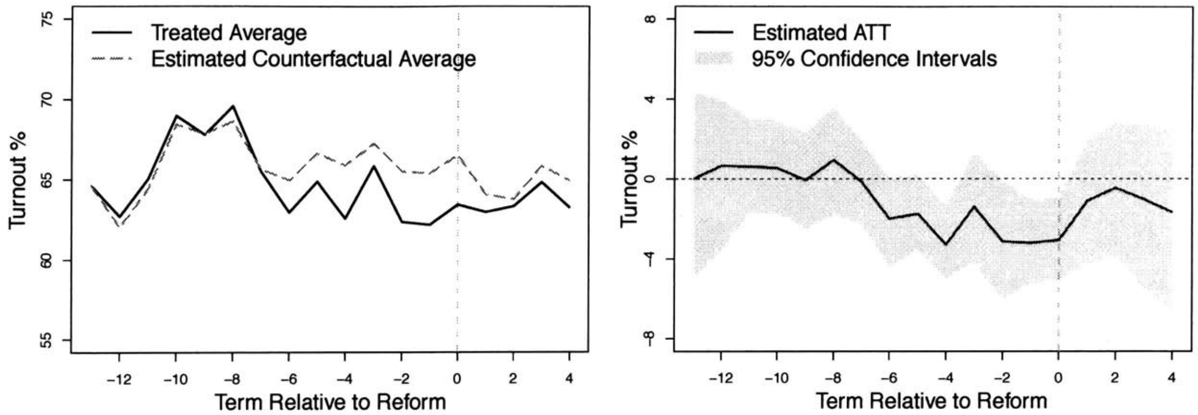
treated states with a set of dummy variables indicating the number for terms relative to EDR reform.[10] The result is visualized in the upper panel of Figure 6. Figure (a) shows average actual turnout rates (black solid line) and average predicted turnout rates in the absence of EDR laws (red dashed line) for the 9 treated states; both averages are taken based on the number of terms since (or before) EDR laws first took effect. Figure (b) shows the gap between the two lines, or the estimated ATT. It is clear from both figures that the "parallel trends" assumption fails: the average predicted turnout deviates from the average actual turnout in pre-treatment periods and the difference is statistically significant from zero at the 5 percent level in multiple periods.

Next, I apply the GSC method to the same dataset. Table 8 columns (3) and (4) summarize the result.[11] Again, both specifications impose additive state and year fixed

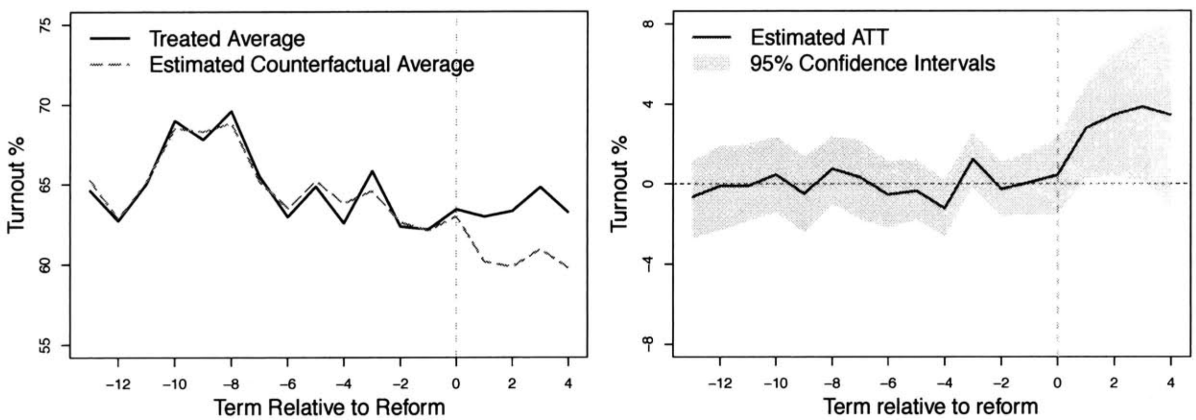
[10]For example, Maine's EDR law first took effect in the 1976 presidential election; hence, 1976 is marked as term 1 while 1972 and 1980 are marked as term 0 and term 2, respectively.

[11]Note that although the estimated ATT of EDR on voter turnout is presented in the same row as the

Figure 6. THE EFFECT OF EDR ON TURNOUT: MAIN RESULTS




(a) Difference-in-differences




(b) Generalized synthetic control

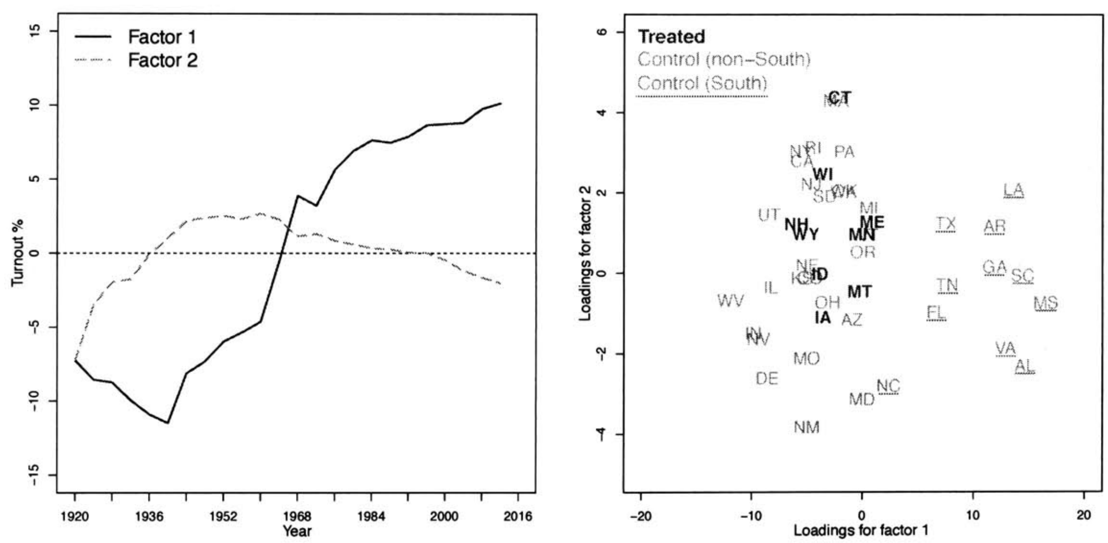
effects. In column (3), no covariates are included, while in column (4), mail-in and motor voter registration are controlled for (assuming that they have constant effects on turnout). With both specifications, the cross-validation scheme finds two unobserved factors to be important and after conditioning on both the factors and additive fixed effects, the estimated ATT based on the GSC method is around 5 percent with a standard error of 2.3 percent.[12] This means that EDR laws are associated with a statistically significant

coefficient of EDR using the DID model, the GSC method does not assume the treatment effect to be constant. In fact, it allows the treatment effect to be different both across states and over time. Predicted counterfactuals and individual treatment effect for each of the 9 treated states are shown in Figure 8 in the Appendix.

[12]The results are similar if additive state and year fixed effects are not directly imposed, though not surprisingly, the algorithm includes an additional factor.

increase in voter turnout, consistent with previous OLS results based on individual-level data. The lower panel of Figure 6 shows the dynamics of the estimated ATT. Again, in the left figure, averages are taken after the actual and predicted turnout rates are re-aligned to the timing of the reform. With the GSC method, the average actual turnout and average predicted turnout match well in pre-treatment periods and diverge after EDR laws took effect. The right figure shows that the gaps between the two lines are not significantly different from zero in pre-treatment periods. The effect takes off right after the adoption of EDR.[13]

Figure 7. THE EFFECT OF EDR ON TURNOUT: FACTORS AND LOADINGS




(a) Factors (b) Loadings

Figure 7 presents the estimated factors and factor loadings produced by the GSC method.[14] Figure 7(a) depicts the two estimated factors. The x-axis is year and the y-axis is the magnitude of factors (re-scaled by the square root of their corresponding eigenvalues to demonstrate their relative importance). Figure (b) shows the estimated factors loadings for each treated (blue, bold) and control (gray) units, with x- and y-axes indicating the magnitude of the loadings for the first and second factors, respectively.

[13]Although it is not guaranteed, this is not surprising since the GSC method uses information of all past outcomes and minimizes gaps between actual and predicted turnout rates in pre-treatment periods.
[14]The results are essentially the same with or without controlling for the other two registration reforms.

Bearing in mind the caveat that estimated factors may not be directly interpretable because they are, at best, linear transformations of the true factors, the estimated factors shown in this figure are meaningful. The first factor captures the sharp increase in turnout in the southern states because of the 1965 Voting Rights Act that removed Jim Crow laws, such as poll taxes or literacy tests, that suppressed turnout. As shown in the right figure, the top 11 states that have the largest loadings on the first factor are exactly the 11 southern states (which were previously in the confederacy).[15] The labels of these states are underlined in Figure 7(b). The second factor, which is set to be orthogonal to the first one, is less interpretable. However, its non-negligible magnitude indicates a strong downward trend in voter turnout in many states in recent years. Another reassuring finding shown by Figure 7(b) is that the estimated factor loadings of the 9 treated units mostly lie in the convex hull of those of the control units, which indicates that the treated counterfactuals are produced mostly by more reliable interpolations instead of extrapolations.

Finally, I investigate the heterogeneous treatment effects of EDR laws. Previous studies have suggested that the motivations behind enacting these laws are vastly different between the early adoptors and later ones. For example, Maine, Minnesota, and Wisconsin, which established the EDR in mid-1970s, did so because officials in these states sincerely wanted the turnout rates to be higher, while the "reluctant adoptors," including Idaho, New Hampshire, and Wyoming, introduced the EDR as a means to avoid the NVRA because officials viewed the NVRA as "a more costly and potentially chaotic system" (Hanmer 2009). Because of the different motivations and other reasons, I may expect the treatment effect of EDR laws to be different in states that adopted them in different times.

The estimation of heterogeneous treatment effects is embedded in the GSC method

[15]Although I can control for indicators of Jim Crow laws in the model, such indicators may not be able to capture the heterogeneous impacts of these laws on voter turnout in each state.

Table 9. THE EFFECT OF EDR ON VOTER TURNOUT: THREE WAVES

| *Outcome variable* | Voter Turnout % | | |
|---|---|---|---|
| | 1st Wave | 2nd Wave | 3rd Wave |
| | (1) | (2) | (3) |
| Election Day Registration | 7.26 | 2.16 | -1.14 |
| | (3.46) | (2.72) | (2.97) |
| Mail-in and motor voter registration | x | x | x |
| State fixed effects | x | x | x |
| Year fixed effects | x | x | x |
| Unobserved factors | 2 | 2 | 2 |
| Observations | 1,128 | 1,128 | 1,128 |
| Treated states | 3 | 3 | 3 |
| | (ME, MN, WI) | (ID, NH, WY) | (NT, IA, CT) |
| Control states | 38 | 38 | 38 |

**Note:** Standard errors are based on parametric bootstraps (blocked at the state level) of 1,000 times.

since it gives individual treatment effects for all treated units in a single run. Table 9 summarizes the ATTs of EDR on voter turnout among the three waves of EDR adoptors. Again, additive state and year fixed effects, as well as indicators of two other registration systems, are controlled for. Table 9 shows that EDR laws have a large and positive effect on the early adoptors (the estimate is 7.26 percent with a standard error of 3.46 percent) while EDR laws were found to have no statistically significant impact on the other six states.[16] Such differential outcomes can be due to two reasons. First, the NVRA of 1993 substantially reduced the cost of registration: since almost everyone who has some intention to vote is a registrant after the NVRA was enacted, "there is now little room for enhancing turnout further by making registration easier" (Highton 2004). Second, because states having a strong "participatory culture" is more likely to be selected into

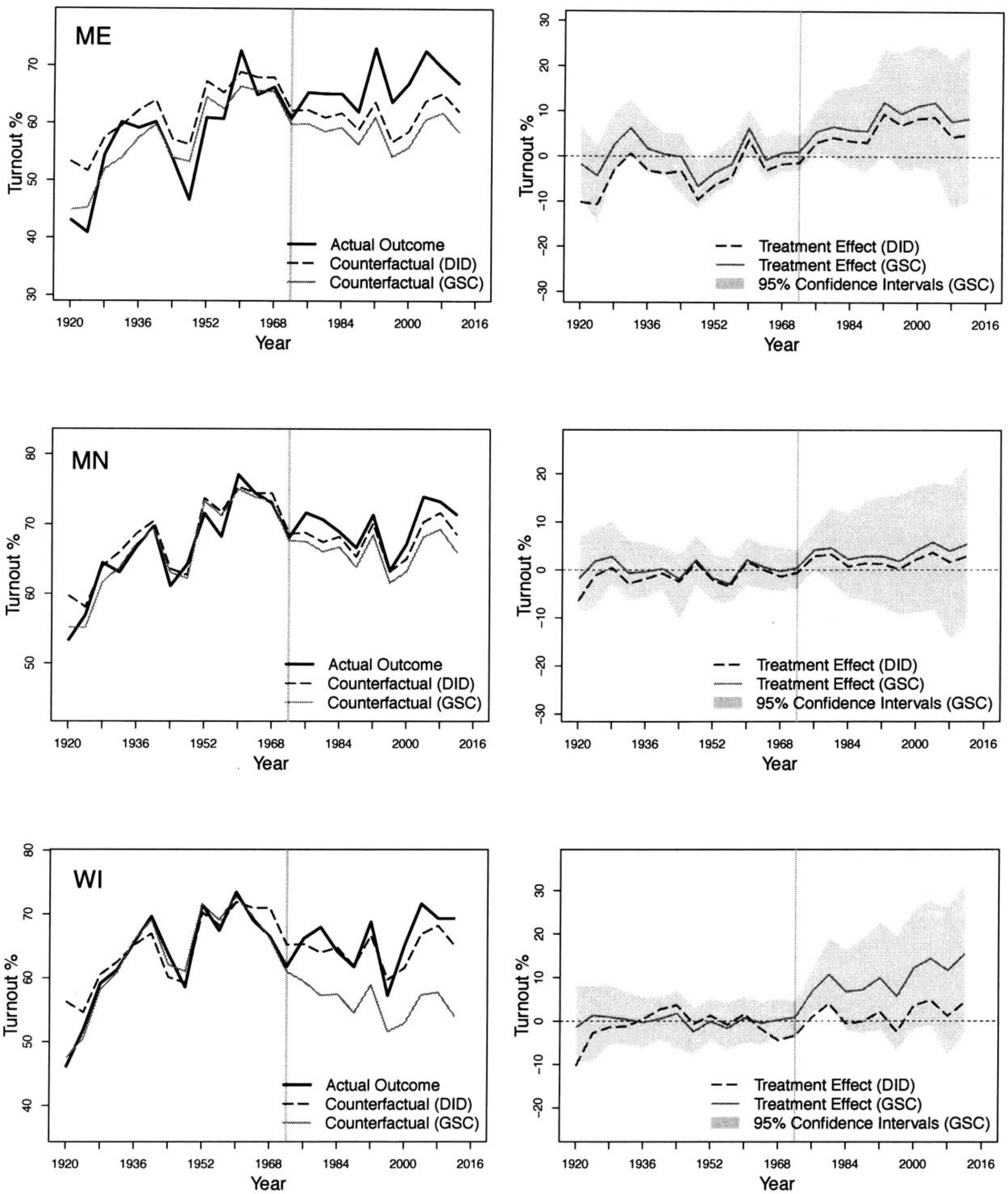
[16]Figure 8 in the Appendix shows that the treatment effects are large and positive for all three early adopting states, Maine, Minnesota, and Wisconsin. Using a fuzzy regression discontinuity design, Keele and Minozzi (2013) show that EDR has almost no effect on the turnout in Wisconsin. The discrepancy with this paper could be mainly due to the difference in the estimands. Two biggest cities in Wisconsin, Milwaukee and Madison constitute a major part of Wisconsin's constituency but have neglectable influence to their local estimates.

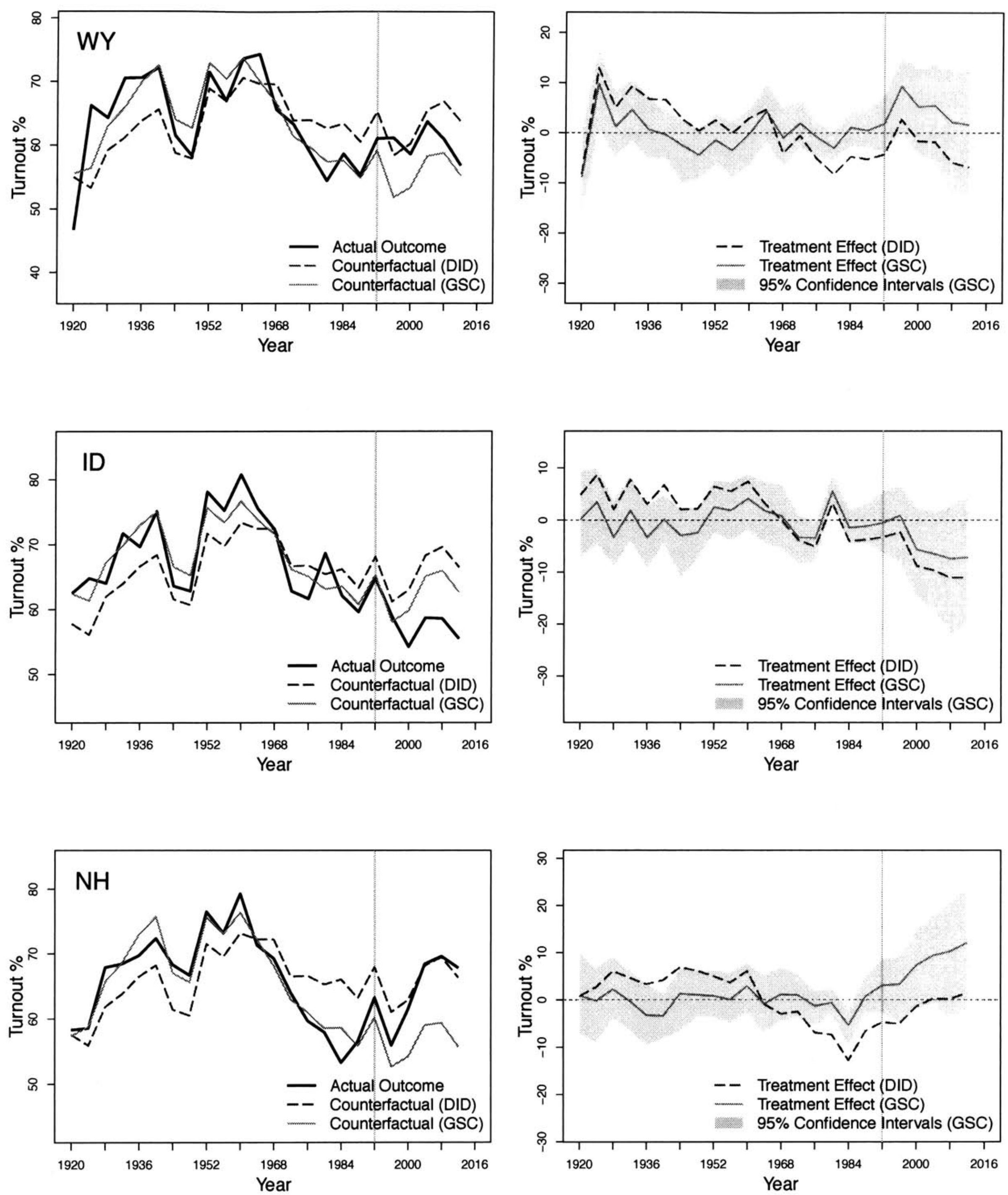
an EDR system in earlier years, costly registration, as a binding constraint in these states, may not be a first-order issue in a state where many eligible voters have low incentives to vote in the first place. It is also possible that voters in early adopting states formed a habit to vote in the days when the demand for participation was high (Hanmer 2009).

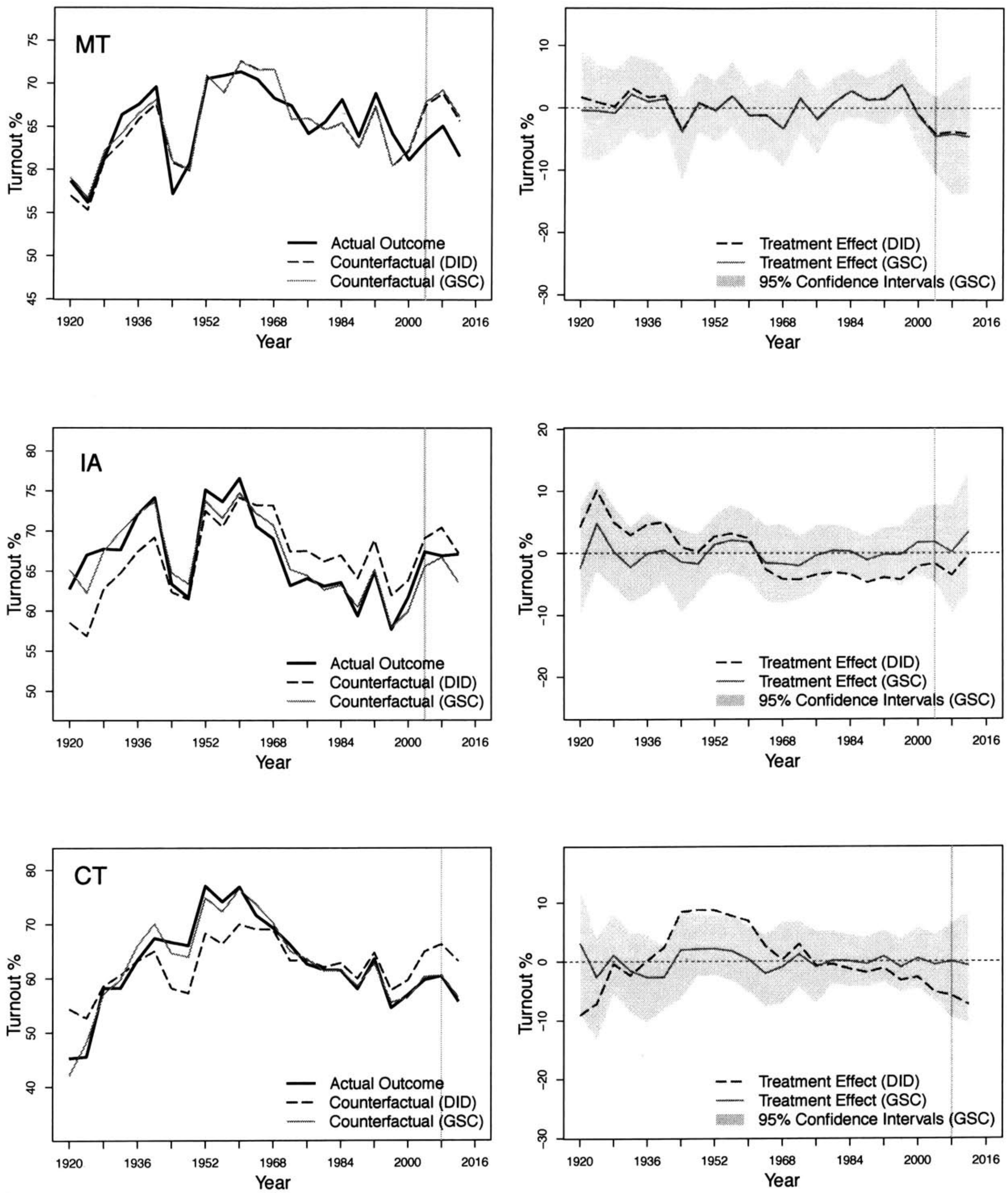
In summary, using the GSC method, I find that EDR laws increased turnout in early adopting states, including Maine, Minnesota, and Wisconsin, but not in states that introduced EDR as a strategy to opt out the NVRA or enacted EDR laws in recent years. These results are broadly consistent with evidence provided by a large literature based on individual-level cross-sectional data. They are also more credible than results from conventional fixed effects models when the "parallel trends" assumption appears to fail.

# Appendix

Figure 8. The Effect of EDR on Turnout: Individual Cases




WY
ID
NH
Turnout %
Year
Actual Outcome
Counterfactual (DID)
Counterfactual (GSC)
Treatment Effect (DID)
Treatment Effect (GSC)
95% Confidence Intervals (GSC)
1920
1936
1952
1968
1984
2000
2016




**Note:** These figures show the fitted values/counterfactuals and estimated individual treatment effect produced by both the DID and GSC estimators for each of the 9 states that enacted EDR laws before the 2012 presidential election. The shades represent the 95% confidence intervals for the treatment effects produced by the GSC method.

## Additional Simulation Results

Table 10. Finite Sample Properties

| | | $N_{tr}=1$ | | | $N_{tr}=5$ | | | $N_{tr}=20$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $T_0$ | $N_{co}$ | Bias | SD | RMSE | Bias | SD | RMSE | Bias | SD | RMSE |
| 5 | 40 | -0.094 | 2.129 | 1.879 | -0.053 | 1.169 | 1.081 | -0.074 | 0.881 | 0.854 |
| 5 | 80 | -0.048 | 1.985 | 1.727 | -0.035 | 0.995 | 0.885 | -0.025 | 0.671 | 0.631 |
| 5 | 120 | -0.024 | 1.942 | 1.665 | -0.020 | 0.936 | 0.825 | -0.014 | 0.590 | 0.546 |
| 5 | 200 | -0.021 | 1.917 | 1.635 | -0.010 | 0.901 | 0.782 | -0.013 | 0.529 | 0.478 |
| 15 | 40 | 0.006 | 1.623 | 1.279 | 0.013 | 0.847 | 0.722 | -0.003 | 0.610 | 0.568 |
| 15 | 80 | 0.004 | 1.555 | 1.203 | 0.007 | 0.769 | 0.624 | 0.002 | 0.486 | 0.434 |
| 15 | 120 | 0.004 | 1.553 | 1.178 | 0.004 | 0.741 | 0.590 | -0.001 | 0.445 | 0.385 |
| 15 | 200 | -0.003 | 1.536 | 1.166 | 0.001 | 0.715 | 0.556 | 0.002 | 0.406 | 0.336 |
| 30 | 40 | 0.005 | 1.556 | 1.187 | -0.012 | 0.813 | 0.674 | -0.009 | 0.563 | 0.518 |
| 30 | 80 | -0.022 | 1.513 | 1.129 | -0.006 | 0.734 | 0.580 | -0.003 | 0.460 | 0.404 |
| 30 | 120 | -0.006 | 1.487 | 1.114 | -0.004 | 0.705 | 0.546 | -0.006 | 0.422 | 0.355 |
| 30 | 200 | 0.000 | 1.470 | 1.089 | -0.007 | 0.681 | 0.521 | -0.004 | 0.382 | 0.314 |
| 50 | 40 | -0.025 | 1.521 | 1.145 | -0.015 | 0.782 | 0.643 | -0.017 | 0.551 | 0.502 |
| 50 | 80 | -0.012 | 1.475 | 1.090 | -0.002 | 0.715 | 0.558 | -0.011 | 0.445 | 0.386 |
| 50 | 120 | -0.006 | 1.459 | 1.068 | -0.006 | 0.695 | 0.525 | -0.009 | 0.409 | 0.341 |
| 50 | 200 | -0.002 | 1.451 | 1.051 | 0.003 | 0.679 | 0.502 | -0.004 | 0.376 | 0.304 |

**Note:** This table shows the bias, standard deviation, and root mean square error of $\widehat{ATT}_{T_0+5}$ with different combinations of $N_{tr}$, $N_{co}$, and $T_0$. Each set of numbers is based on 20,000 simulated samples. The data generating process (DGP) is as follows:

$$Y_{it} = \delta_{it}D_{it} + x_{it,1}\cdot 1 + x_{it,2}\cdot 3 + \lambda_{i1}f_{1t} + \lambda_{i2}f_{2t} + \alpha_i + \xi_t + 5 + \varepsilon_{it}.$$

(1) The treatment indicator $D_{it}$ equal 1 when $i \in \mathcal{T}$ and $t > T_0$ and 0 otherwise. Heterogenous treatment effects are generated by $\delta_{it,t>T_0} = \bar{\delta}_t + e_{it}$, in which $e_{it}$ is i.i.d. N(0,1) and $[\bar{\delta}_{T_0+1}, \bar{\delta}_{T_0+1}, \cdots, \bar{\delta}_{T_0+10}] = [1, 2, \cdots, 10]$. Hence, $\mathbb{E}[ATT_{T_0+5}] = 5$. (2) The regressors are generated by: $x_{it,k} = 1 + \frac{1}{2}\sum_{s=1}^{2}\lambda_{is}f_{st} + \frac{1}{4}\lambda_{i1} + \frac{1}{4}\lambda_{i2} + \frac{1}{4}f_{1t} + \frac{1}{4}f_{2t} + \eta_{it,k}$, $k = 1, 2$ in which $\eta_{it,k}$ is i.i.d. $N(0,1)$. (3) The factors $f_{st}, s = 1, 2$ and time fixed effects $\xi_t$, are i.i.d. $N(0,1)$ for all $t$ and are drawn only once. (4) The factor loadings $\lambda_{is}, s = 1, 2$ and unit fixed effects $\alpha_i$ are i.i.d. $U[-\sqrt{3}, \sqrt{3}]$ for $i \in \mathcal{C}$ and i.i.d. $U[0, 2\sqrt{3}]$ for $i \in \mathcal{T}$, which produces positive correlations between the treatment indicator, the factor loadings, and the regressors. (5) The error term $\varepsilon_{it}$ is i.i.d. $N(0,1)$.

Table 11. EXTENSION: IMPOSING ADDITIVE FIXED EFFECTS

| | | | | w/o additive FEs | | | w/ additive unit FEs | | | w/ additive two-way FEs | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T_0$ | $N_{co}$ | $N_{tr}$ | $r$ | Bias | SD | RMSE | Bias | SD | RMSE | Bias | SD | RMSE |
| | | | | | | *Panel A: DGP w/o additive FEs* | | | | | | |
| 15 | 10 | 40 | 2 | 0.023 | 0.653 | 0.571 | 0.013 | 0.675 | 0.593 | 0.012 | 0.705 | 0.628 |
| 15 | 10 | 80 | 2 | 0.010 | 0.560 | 0.464 | 0.004 | 0.571 | 0.478 | 0.004 | 0.585 | 0.496 |
| 15 | 10 | 120 | 2 | 0.010 | 0.538 | 0.433 | 0.006 | 0.547 | 0.446 | 0.007 | 0.558 | 0.457 |
| | | | | | | *Panel B: DGP w/ additive FEs* | | | | | | |
| 15 | 10 | 40 | 2 | -0.003 | 0.747 | 0.679 | 0.008 | 0.698 | 0.625 | 0.006 | 0.696 | 0.623 |
| 15 | 10 | 80 | 2 | -0.001 | 0.636 | 0.554 | 0.017 | 0.597 | 0.511 | 0.016 | 0.594 | 0.509 |
| 15 | 10 | 120 | 2 | 0.012 | 0.594 | 0.507 | 0.016 | 0.558 | 0.464 | 0.016 | 0.557 | 0.463 |

**Note:** This table compares performances of the generalized synthetic control estimator for $ATT_{T_0+5}$ when additive fixed effects are not imposed in the model and when they are not under the circumstance that additive fixed effects are not present (Panel A) and present (Panel B). Each set of numbers is based on 5,000 simulated samples. The data generating process (DGP) is as follows:

$$Y_{it} = \delta_{it}D_{it} + x_{it,1}\cdot 1 + x_{it,2}\cdot 3 + \sum_{s=1}^{r}\lambda_{is}f_{st} + 5 + \varepsilon_{it} \quad \text{for Panel A, and}$$

$$Y_{it} = \delta_{it}D_{it} + x_{it,1}\cdot 1 + x_{it,2}\cdot 3 + \sum_{s=1}^{r}\lambda_{is}f_{st} + \alpha_i + \xi_t + 5 + \varepsilon_{it} \quad \text{for Panel B.}$$

(1) The treatment indicator $D_{it}$ equal 1 when $i \in \mathcal{T}$ and $t > T_0$ and 0 otherwise. Heterogenous treatment effects are generated by $\delta_{it,t>T_0} = \bar{\delta}_t + e_{it}$, in which $e_{it}$ is i.i.d. N(0,1) and $[\bar{\delta}_{T_0+1}, \bar{\delta}_{T_0+1}, \cdots, \bar{\delta}_{T_0+10}] = [1, 2, \cdots, 10]$. Hence, $\mathbb{E}[ATT_{T_0+5}] = 5$. (2) The regressors are generated by: $x_{it,k} = 1 + \frac{1}{2}\sum_{s=1}^{2}\lambda_{is}f_{st} + \frac{1}{4}\lambda_{i1} + \frac{1}{4}\lambda_{i2} + \frac{1}{4}f_{1t} + \frac{1}{4}f_{2t} + \eta_{it,k}$, $k = 1, 2$ in which $\eta_{it,k}$ is i.i.d. $N(0,1)$. (3) The factors $f_{st}, s = 1, 2, \cdots, r$ and time fixed effects $\xi_t$, are i.i.d. $N(0,1)$ for all $t$ and are drawn only once. (4) The factor loadings $\lambda_{is}, s = 1, 2, \cdots, r$ and unit fixed effects $\alpha_i$ are i.i.d. $U[-\sqrt{3}, \sqrt{3}]$ for $i \in \mathcal{C}$ and i.i.d. $U[0, 2\sqrt{3}]$ for $i \in \mathcal{T}$, which produces positive correlations between the treatment indicator, the factor loadings, and the regressors. (5) The error term $\varepsilon_{it}$ is i.i.d. $N(0,1)$.

Table 12. UNIFORM FACTORS

| | | $N_{tr}=1$ | | | $N_{tr}=5$ | | | $N_{tr}=20$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $T_0$ | $N_{co}$ | Bias | SD | RMSE | Bias | SD | RMSE | Bias | SD | RMSE |
| 5 | 40 | 0.002 | 1.753 | 1.442 | 0.021 | 0.932 | 0.818 | 0.008 | 0.685 | 0.644 |
| 5 | 80 | 0.017 | 1.679 | 1.346 | 0.011 | 0.828 | 0.698 | 0.018 | 0.538 | 0.488 |
| 5 | 120 | 0.011 | 1.669 | 1.332 | 0.004 | 0.795 | 0.657 | 0.005 | 0.481 | 0.428 |
| 5 | 200 | -0.002 | 1.652 | 1.304 | 0.003 | 0.760 | 0.617 | 0.007 | 0.436 | 0.375 |
| 15 | 40 | 0.081 | 1.564 | 1.220 | 0.069 | 0.817 | 0.695 | 0.069 | 0.578 | 0.536 |
| 15 | 80 | 0.040 | 1.534 | 1.154 | 0.040 | 0.744 | 0.598 | 0.038 | 0.472 | 0.416 |
| 15 | 120 | 0.006 | 1.519 | 1.129 | 0.018 | 0.716 | 0.561 | 0.028 | 0.428 | 0.365 |
| 15 | 200 | -0.000 | 1.496 | 1.110 | 0.015 | 0.700 | 0.537 | 0.014 | 0.391 | 0.323 |
| 30 | 40 | 0.050 | 1.525 | 1.164 | 0.052 | 0.790 | 0.655 | 0.058 | 0.547 | 0.501 |
| 30 | 80 | 0.013 | 1.484 | 1.104 | 0.022 | 0.721 | 0.563 | 0.027 | 0.451 | 0.395 |
| 30 | 120 | 0.015 | 1.462 | 1.073 | 0.018 | 0.694 | 0.532 | 0.015 | 0.414 | 0.346 |
| 30 | 200 | 0.001 | 1.459 | 1.062 | 0.013 | 0.673 | 0.502 | 0.010 | 0.375 | 0.303 |
| 50 | 40 | 0.068 | 1.527 | 1.137 | 0.073 | 0.788 | 0.653 | 0.069 | 0.547 | 0.504 |
| 50 | 80 | 0.026 | 1.473 | 1.088 | 0.030 | 0.720 | 0.558 | 0.037 | 0.446 | 0.386 |
| 50 | 120 | 0.021 | 1.451 | 1.064 | 0.021 | 0.690 | 0.525 | 0.020 | 0.407 | 0.342 |
| 50 | 200 | 0.017 | 1.438 | 1.048 | 0.010 | 0.668 | 0.497 | 0.012 | 0.371 | 0.299 |

**Note:** This table shows the bias, standard deviation, and root mean square error of $\widehat{ATT}_{T_0+5}$ with different combinations of $N_{tr}$, $N_{co}$, and $T_0$ when time-varying factors (including time fixed effects) are drawn from a uniform distribution. Each set of numbers is based on 20,000 simulated samples. The data generating process (DGP) is as follows:

$$Y_{it} = \delta_{it}D_{it} + x_{it,1}\cdot 1 + x_{it,2}\cdot 3 + \lambda_{i1}f_{1t} + \lambda_{i2}f_{2t} + \alpha_i + \xi_t + 5 + \varepsilon_{it}.$$

(1) The treatment indicator $D_{it}$ equal 1 when $i \in \mathcal{T}$ and $t > T_0$ and 0 otherwise. Heterogenous treatment effects are generated by $\delta_{it,t>T_0} = \bar{\delta}_t + e_{it}$, in which $e_{it}$ is i.i.d. N(0,1) and $[\bar{\delta}_{T_0+1}, \bar{\delta}_{T_0+1}, \cdots, \bar{\delta}_{T_0+10}] = [1, 2, \cdots, 10]$. Hence, $\mathbb{E}[ATT_{T_0+5}] = 5$. (2) The regressors are generated by: $x_{it,k} = 1 + \frac{1}{2}\sum_{s=1}^{2}\lambda_{is}f_{st} + \frac{1}{4}\lambda_{i1} + \frac{1}{4}\lambda_{i2} + \frac{1}{4}f_{1t} + \frac{1}{4}f_{2t} + \eta_{it,k}$, $k = 1, 2$ in which $\eta_{it,k}$ is i.i.d. $N(0,1)$. (3) The factors $f_{st}, s = 1, 2$ and time fixed effects $\xi_t$, are i.i.d. $U[-\sqrt{3}, \sqrt{3}]$ for all $t$ and are drawn only once. (4) The factor loadings $\lambda_{is}, s = 1, 2$ and unit fixed effects $\alpha_i$ are i.i.d. $U[-\sqrt{3}, \sqrt{3}]$ for $i \in \mathcal{C}$ and i.i.d. $U[0, 2\sqrt{3}]$ for $i \in \mathcal{T}$, which produces positive correlations between the treatment indicator, the factor loadings, and the regressors. (5) The error term $\varepsilon_{it}$ is i.i.d. $N(0,1)$.

# Bibliography


Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490):493–505.

Bai, Jushan. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77:1229–1279.

Blackwell, Matthew and Adam Glynn. 2015. "How to Make Causal Inferences with Time-Series Cross-Sectional Data." Mimeo, Harvard Uniersity.

Brians, Craig Leonard and Bernard Grofman. 2001. "Election Day Registration's Effect on US Voter Turnout." *Social Science Quarterly* 82(1):170–183.

Burden, Barry C., David T. Canon, Kenneth R. Mayer and Donald P. Moynihan. 2009. The Effects and Costs of Early Voting, Election Day Registration, and Same Day Registration in the 2008 Elections. In *Annual meeting of the American Political Science Association, Toronto.* Citeseer.

Burnham, Walter Dean. 1980. The Appearance and Disappearance of the American Voter. In *Electoral Participation: A Comparative Analysis*, ed. Richard Rose. Beverly Hills, CA: Sage Publications.

Cain, Bruce E., Todd Donovan and Caroline J. Tolbert. 2011. *Democracy in the States: Experiments in Election Reform.* Brookings Institution Press.

Fenster, Mark J. 1994. "The Impact of Allowing Day of Registration Voting On Turnout in US Elections From 1960 To 1992 A Research Note." *American Politics Research* 22(1):74–87.

Gobillon, Laurent and Thierry Magnac. 2013. "Regional Policy Evaluation: Interactive Fixed Eects and Synthetic Controls." Mimeo, Paris School of Economics.

Hanmer, Michael J. 2009. *Discount Voting: Voter Registration Reforms and their Effects.* Cambridge University Press.

Highton, Benjamin. 1997. "Easy Registration and Voter Turnout." *The Journal of Politics* 59(2):565–575.

Highton, Benjamin. 2004. "Voter Registration and Turnout in the United States." *Perspectives on Politics* 2(3):507–515.

Huang, Chi and Todd G. Shields. 2000. "Interpretation of Interaction Effects in Logit and Probit Analyses Reconsidering the Relationship Between Registration Laws, Education, and Voter Turnout." *American Politics Research* 28(1):80–95.

Keele, L. and W. Minozzi. 2013. "How Much Is Minnesota Like Wisconsin? Assumptions and Counterfactuals in Causal Inference with Observational Data." *Political Analysis* 21(2):193–216.

King, James D. and Rodney A. Wambeam. 1995. "Impact of Election Day Registration on Voter Turnout: A Quasi-experimental Analysis." *Policy Studies Review* 14(3):263–278.

Knack, Stephen. 2001. "Election-day Registration The Second Wave." *American Politics Research* 29(1):65–78.

Knack, Stephen and James White. 2000. "Election-day Registration and Turnout Inequality." *Political Behavior* 22(1):29–44.

Mitchell, Glenn E. and Christopher Wlezien. 1995. "The Impact of Legal Constraints on Voter Registration, Turnout, and the Composition of the American Electorate." *Political behavior* 17(2):179–202.

Neiheisel, J. R. and B. C. Burden. 2012. "The Impact of Election Day Registration on Voter Turnout and Election Outcomes." *American Politics Research* 40(4):636–664.

Pang, Xun. 2010. "Modeling Heterogeneity and Serial Correlation in Binary Time-Series Cross-sectional Data: A Bayesian Multilevel Model with AR(p) Errors." *Political Analysis* (4):470–498.

Pang, Xun. 2014. "Varying Responses to Common Shocks and Complex Cross-Sectional Dependence: Dynamic Multilevel Modeling with Multifactor Error Structures for Time-Series Cross-Sectional Data." *Political Analysis* 22(4):464–496.

Park, Jong Hee. 2010. "Structural Change in US Presidents' Use of Force." *American Journal of Political Science* 54(3):766–782.

Park, Jong Hee. 2012. "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models." *American Journal of Political Science* 56(4):1040–1054.

Rhine, Staci L. 1992. "An Analysis of the Impact of Registration Factors on Turnout in 1992." *Political Behavior* 18(2):171–185.

Springer, Melanie Jean. 2014. *How the States Shaped the Nation: American Electoral Institutions and Voter Turnout, 1920-2000.* University of Chicago Press.

Teixeira, Ruy A. 2011. *The Disappearing American Voter.* Brookings Institution Press.

Timpone, Richard J. 1998. "Structure, Behavior, and Voter Turnout in the United States." *The American Political Science Review* 92(1):145.

Timpone, Richard J. 2002. "Estimating Aggregate Policy Reform Effects: New Baselines for Registration, Participation, and Representation." *Political Analysis* 10(2):154–177.

Wolfinger, Raymond E. and Steven J. Rosenstone. 1980. *Who Votes?* New Haven, CT: Yale University Press.

# Chapter 3

*This chapter is coauthored with Yang Yao, Professor at National School of Development at Peking University. It was published in American Political Science Review Volume 109, Issue 2, May 2015, titled "Informal Institutions, Collective Action, and Public Investment in Rural China." This thesis reprints the chapter with permission from Cambridge University Press.*

Do informal institutions promote good governance in localities where formal democratic and bureaucratic institutions are weak? Or do they prevent local governments from functioning properly? Scholars find that in Latin America, Eastern Europe, and Central Asia informal institutions often breed clientelism, corruption, and mafia activities (e.g., O'Donnell 1996; Böröcz 2000; Collins 2003) and cause citizens to be excluded from the state's public services (Narayan 1999). However, Sklar (2004) suggests that traditional institutions in Uganda and Nigeria improve government performance and maintain regime stability. Tsai (2007) shows that in the context of rural China, solidary groups, such as temple associations and village-wide lineage groups, hold government officials accountable and motivate them to provide more public goods.

Although a universal answer to the question may not exist, a clearer understanding of the role of informal institutions in specific social contexts deepens our knowledge of what determines good local governance. However, researchers on informal institutions often face challenges in defining and measuring informal institutions and in identifying their causal effects on governance outcomes (Helmke and Levitsky 2004). We attempt to address those challenges. We follow Helmke and Levitsky (2004) and define informal institutions as rules and norms that are created and enforced by social groups rather than the state. In this paper, we specifically focus on the set of informal institutions that could affect local public goods provision.

Public goods provision in environments of weak formal institutions faces two fundamental problems: (1) to convince community members, who are often in poor living conditions and have tight budget constraints, to contribute to public goods expenditure and (2) to motivate local leaders to initiate necessary public projects, while preventing their moral hazard behavior, such as embezzlement and corruption, during the process of providing public goods. The first problem is essentially a collective action problem, while the second one is about local government accountability. If informal institutions are to promote local public goods provision, it is likely they either help solve the collective action problem among community members, or hold local officials accountable, or both.

Using a panel dataset of 220 Chinese villages from 1986 to 2005, we study the effect of informal institutions embedded in large and organized family clans on public goods provision and investigate the collective action and accountability mechanisms. Family clans are organized based on shared patrilineal ancestors and are regarded as the most important social groups in Chinese villages (e.g., Fei 1946; Freedman 1958; Watson 1982, Duara 1988). Informal institutions of large clans are rules created and enforced by clans and often respected by villagers both inside and outside clans. Large clans may have disproportionate advantage over small clans because they have deeper historical roots in the village and are often better organized. As a result, they may promote local public goods provision by either helping the officials coordinate collective action or by holding them accountable.

The exact outcome variable in our analysis is the amount of public investment the village committee spent each year in the period of 1986-2005 after village elections were introduced. We focus on the post-election period because it is the period during which we have complete data on elected village chairpersons (VCs). The key independent variables are binary indicators of whether a village leader, such as an elected VC or an appointed village party secretary (VPS), came from the village's largest or second-largest clan. Our theoretical premise is that leaders from these two clans have access to richer and stronger informal institutions than leaders from small clans. Specifically, there are two possible channels that a village leader might be affected by his own clan when he attempts to initiate a public investment project: (1) he could get help from his clan and use the clan's social power to mobilize resources from villagers and (2) he might be morally bounded by the rules of his clan such that he would make good use of existing resources. In either case, large clans dominate small clans in terms of public goods provision because of their advantageous positions in the village. To further illustrate that it is the rules and norms of a large clan, rather than the number of its members, that matters, we use the information on whether a large clan kept records of family trees and whether it maintained a lineage hall since the beginning of the observed data period to create indicators of clan cohesiveness and investigate how the effect of informal institutions changes due to the changes in the level of cohesiveness.

Because variations of informal institutions in existing quantitative studies are usually cross-sectional, it is often difficult to identify the effect of informal institutions on governance outcomes. One can imagine a situation in which associational activities flourish in places with better infrastructure or rapidly increasing living standards. The positive correlation between associational activities and public goods provision does not necessarily imply that the former, which are sometimes used as proxies for informal institutions, cause the latter. Controlling for time-invariant heterogeneity would alleviate this concern of omitted variables to a great extent. In this paper, we exploit the advantage of the panel data structure and only look at the within-village changes of public goods expenditure due to within-village changes of informal institutions associated with village leaders. To the extent that informal institutions affect local governance, we would then expect to observe systematically different policy outcomes produced by villager leaders of different clans within the same village.

Our research design, therefore, is to compare the level of public goods expenditure during the terms of village leaders who came from the village's two largest clans and the terms of the others within each village (focussing on the largest clan gives qualitatively the same results). We primarily focus on VCs instead of VPSs because we have more complete data on the former than the latter. To address the concern that electoral outcomes might be endogenous to public goods expenditure—for example, villagers expect leaders from large clans to provide more public goods and, therefore, elect them into office—we conduct a regression discontinuity analysis based on elected VCs as a robustness check for our main results.

Setting the study in the context of rural China has several advantages. First, because of the large scale of the country, there is enough variation in the lineage composition of a village. Forms, origins, and functions of the institutions associated with lineage groups are relatively well understood by scholars (e.g., Fei 1946; Freedman 1958; Watson 1975; Wang 1996; Duara 1988; Tsai 2007), which makes it considerably easier to understand how such informal institutions work than in other less researched contexts. Second, Chinese villages are relatively homogeneous in other aspects and subject to similar social and political shocks at the provincial or national level. This aspect makes them better comparison groups of each other than nations in cross-country studies. Third, Chinese villages were largely autonomous in terms of determining and financing public goods in the period of our study. Fourth, the introduction of village elections in the mid-1980s offers a rare chance to examine the influence of both formal and informal institutions.

Our empirical analysis shows that during the terms of VCs of the two largest clans (hereafter, VCs of large clans), the amount of village public investment increased by more than 35 percent on average as compared with the amount during the terms of other VCs. A VPS of the two largest clans (VPS of large clans) also increased the average level of public investment considerably. We interpret these results as evidence that the informal institutions of lineage groups, rather than village leaders of a certain kind, led to more public goods expenditure and, presumably, better local governance. We show that the association between VCs of large clans and public goods expenditure is stronger in places where large clans appeared to be more cohesive (i.e., clans that had maintained lineage halls since or before the onset of elections). Combined together, these results indicate that it is informal institutions of the clans from which village leaders originated that drive our main finding.

In addition, we explore two mechanisms through which informal institutions of large clans may facilitate public goods provision: (1) the collective action mechanism and (2) the accountability mechanism. The collective action problem has been a central topic in political science since Olson (1965). Recently, researchers have been focusing on how informal institutions, rather than formal ones, help people overcome the collective action problem. For instance, after observing several long-standing, self-governing common property regimes, Ostrom (1990) argues that informal institutions work through (or are) a set of "self-enforcing rules that each community member commits himself or herself to follow" (p. 99). Banerjee and Iyer (2005) suggest that persistent informal institutions may result in different levels of public expenditure because of the nature of collective action embedded in those institutions. In Africa, Habyarimana et al. (2007, 2009) show that ethnic heterogeneity impedes collective action and the provision of public goods; however, the collective action problem can be alleviated by institutional improvements in monitoring, sanctioning, and enforcement.

In rural China, because village committees often lack measures to enforce levies on villagers, successful collection of levies requires villagers' semi-voluntary compliance. Sheer poverty in the countryside makes that difficult for a VC and his associates.[1] If a large proportion of villagers refuse to pay for a public investment project, village leaders' efforts to provide public goods would be in vain without the help of the upper-level governments. We show that when VCs of large clans were in office, villagers paid more levies to the village committee, and the presence of village public investment projects is highly correlated with extra levies paid by villagers at almost all income percentiles. Our results

[1]In 2005, the median household in our sample lived with an annual budget of 18,507 yuan, or US$1.61 per household member per day (purchasing power not adjusted). In 1986, that number was US$0.44 per household member per day.

indicate that with the help of the informal institutions of large clans, elected VCs were more able to enforce levies on villagers and to mobilize resources needed for providing public goods.[2]

Sklar (2004) and Tsai (2007), among others, emphasize the mechanism of informal accountability. To test this hypothesis, we study the amount of administrative costs during each VC's term. Administrative costs are mostly spent by the VC and his associates for their own consumption. Embezzlement and other forms of corruption may also be covered in this category of village spending. A decline of those costs, therefore, can be seen as a result of improved accountability imposed on the VC. However, we do not find evidence that the amount of administrative cost spent by VCs of large clans was smaller than that by VCs of smaller clans. Although we cannot entirely rule out the accountability mechanism, this piece of evidence suggests that the positive association between VCs of large clans and a higher level of public goods expenditure is unlikely to be a result of large clans' superior ability to monitor the VCs.

We investigate two alternative explanations. First, VCs of large clans may be more competent than others. For instance, Munshi and Rosenzweig (2010) show that elected representatives from large castes in rural India exhibit better observed characteristics, such as higher education, and provide more local public goods for their constituents. Second, gradual improvement of formal institutions, such as electoral rules and procedures, may also contribute to the association between VCs of large clans and public goods provision. As elections become more competitive, electoral outcomes are more likely to reflect the preferences of the constituents. Thus, it is possible that the probability of VCs of large clans being elected and increased public goods provision are moving in the same direction. We show that neither of these two possibilities is likely to be driving our results.

[2]Our finding is consistent with Habyarimana et al. (2007)'s finding from experiments in Africa that ethnically more homogeneous communities achieve greater success in collective action because of better communication technology, more transparency, and more cooperative equilibrium strategies.

Apart from the informal institution literature and the literature on collective action and public goods provision, this paper also adds to a large literature on village elections and grassroots politics in China (e.g., O'Brien 1994; Manion 1996, 2006; Shi 1999; Oi 1999; Oi and Rozelle 2000; O'Brien and Li 1999, 2000; Pastor and Tan 2000). More recently, Luo et al. (2007, 2010) find that the introduction of elections increases total public goods expenditure and provision. Shen and Yao (2008) find that elections reduce village income inequality through the public goods channel. Martinez-Bravo et al. (2011) show that the introduction of village elections has shifted accountability of village leaders from the upper-level government towards villagers and worsened the implementation of unpopular policies, such as tax collection and the One Child Policy. This paper, instead of investigating the effect of elections *per se*, uses variations generated by elections to examine the causal effect of informal institutions on governance outcomes. To the best of our knowledge, this paper is also the first to apply a regression discontinuity design to village elections in China.

## Institutional Background

**Lineage groups in Chinese villages.** Lineage groups are one of the most important social organizations in rural China. They are usually organized along the paternal line.[3] Fei (1946) suggests that in imperial times lineage groups served as a link between the imperial ruler and the grassroots and were used by the gentry to preserve the social and political power of their families. Fei finds that through lineage networks the gentry administrated charities and provided local public goods to command the moral height in the villages. Freedman (1958) hypothesizes that lineage groups are substitutive social organizations in places where formal bureaucratic institutions are weak. He finds

[3] Watson (1982) defines a lineage group as "a corporate group which celebrates ritual unity and is based on demonstrated descent from a common ancestor." He distinguishes clans from lineage groups based on membership recruitment. He argues that clans recruit members based on fictionalized descent rather than descent from known ancestors. However, most scholars do not distinguish the two terms. In this paper, we focus on clans that are formed based on known ancestors and use the two terms interchangeably.

that lineage organizations were more developed in southeastern China than in the north because villages in the southeast were farther away from central political control.

After nearly one hundred years of radical social changes, there has been a startling withdrawal of the gentry from the rural political field.[4] However, researchers believe that there is still space for lineage groups to survive and flourish. Reformers and even revolutionaries had to take advantage of existing resources, including traditional institutions, to achieve their objectives (Perry 2002). Lineage groups have proven resilient and, in many places, have survived extreme social and political changes (Wang 1996).[5] Tsai (2007) reports in her 2001 survey that 14 percent of the villages had one or more lineage halls.

Previous research on the Chinese village focuses on ideal types of social organizations, such as village-wide lineage groups. Sub-village lineage groups are thought to be not as effective in exercising social powers (e.g., Freedman 1958; Tsai 2007). However, the introduction of village elections may activate some of the functions of sub-village lineage groups. These groups are often organized around surnames or, when the village has only one surname, *fang* (house), that is, households who share the same grand- or great-grandparents. In the absence of political parties or other modern political organizations, lineage groups can become vehicles for political mobilization. A clan can be as large as 100 households containing more than 400 villagers, although it may only constitute 30 percent of the total village population. Such a group, if well organized, can have a real impact on village governance. In this paper, variation in informal institutions comes from these groups.

As a social organization, family clans in rural China have several features. First, as mentioned above, households within a clan consciously identify themselves as members of

[4]The reason behind the change is complicated. The rise of towns and cities since the beginning of the twentieth century attracted the young and wealthy out of the rural areas. The neighborhood administrative system (*baojia zhi*) in the Republican era and endless social movements after 1949 also contributed to the retreat of clan forces.

[5]This finding is consistent with researchers' finding in Central Asia that clans adapt to resist repressive states (Collins 2004).

a closely-bonded group. They often reside within geographical proximity and frequently interact with one another. A well-organized clan holds annual rituals and ceremonies, such as paying respect to ancestral tablets and offering sacrifices to ancestral spirits (usually at its lineage hall), to reinforce group identity (e.g., Freedman 1967, Tsai 2007). Second, clan members often cooperate with each other to obtain material benefits. Before the communist revolution in 1949, clans in southern China often owned land, which gave a basis for clan members to cooperate economically. In the collectivization period, collective production teams were often organized by clans in the south. Although this happened mostly because of the geographic proximity of clan members' residencies, economic ties within the clan were preserved (Watson 1982). In the reform period, economic cooperation among clan members has shifted to other areas of shared interests. For example, rural entrepreneurs tend to hire relatives in their own firms (Oi 1999: 69). Third, clan members share a sense of obligation to the group. Traditional ethics place a sacred value on loyalties generated by kinship and dense social ties. Moral standing is conferred to members who make contributions or bring material benefits to the group (Madsen 1984). Finally, leaders of the group, usually respected senior male members, come forward to enforce social norms and mitigate conflicts both within and outside the group.[6]

Large clans may have disproportionate advantage in this regard. This is first and foremost related to the fact that they often have strong historical roots in villages. Many villages were founded by the largest clans in the first place. Watson (1982) reports that "it is common to find villages that contain one or two corporate lineages together with four or five loosely-defined surname groups" (p. 608). Small clans often consist of families that migrated into the village at a later stage. Because of that, large clans are usually better organized than small clans. In the past, they were often managed by a group

[6]As in other hierarchical social groups, not all members of a clan enjoy the same social status in the group. One can imagine that if a well-respected member of the group gets elected, he or she can mobilize more group resources than others. Our dataset does not have such information, but we find that the profiles and characteristics of elected VCs remain stable over time.

of senior members led by *zuzhang*, or lineage chief, a position usually inherited by the most powerful family in the clan (often the family of the eldest son of the clans founder). Today, this more formal power structure has vanished. However, senior members still play a significant role. They are responsible for clan rituals and other collective activities (Cohen 1990).

Large clans thus are more likely than small clans to maintain lineage halls, hold clan ceremonies and keep lineage genealogies. This increases their social cohesiveness and members' sense of belonging. In addition, seniors of large clans are more likely to participate in village affairs. Many villages have a seniors' association that is recognized by the government as a vehicle to serve the needs of the senior. However, village leaders often consult the members of the association on important village affairs. Seniors from large clans naturally become the leaders of the association. As a result, their influence can reach beyond their own clans. Their social power originates from both their clans' clout and their reputation of looking after village public interest.

Against this background, village leaders from large clans can have significant advantages over ones from small clans because it is likely that they can only mobilize informal organization resources from their own clans. Informal institutions embedded in large clans can facilitate collective action among villagers through both *persuasion* and *social sanctioning*. When contribution from villagers is needed for a public project, a village leader from a large clan can approach his clans seniors to ask for help. Resorting to their prestigious social status, senior members of the clan are able to persuade villagers both inside and outside the clan to support the village leader's project and to enforce clan rules and norms when it comes to financial contribution.[7] When a non-trivial proportion of the villagers support the project, social pressure forces the rest of the villagers to contribute

[7]The norms are that, for instance, each household should contribute to the public good according to its own economic condition (*liang li er xing*); households who fail to fulfill their duties will be socially sanctioned by the clan. Liu (1959) asserts that such obligations are often specified in the appendices of genealogies.

their fair shares; otherwise, they may face severe social sanctions.[8]

A piece of anecdotal evidence from Zhejiang Province in eastern China illustrates how the collective action mechanism might work:

> "My father used to be the *zuzhang* (lineage chief) of our clan. The village chief (chairman of the village committee) was also a member of the clan. Whenever the village committee had some great undertakings to accomplish, like collecting money for building a road, he came to my father and other seniors of the clan. If the seniors thought the chief's plan could work, they would convene a meeting of household heads, together with the village chief, to convince the villagers to support the project, either by giving money or donating working hours. Since in our village, the majority of households are from the *Fu* family, the meeting is almost like a villagers' assembly. People took it quite seriously. They trusted my father because they thought he's impartial and experienced. The seniors didn't enjoy formal titles, and they didn't take charge of daily matters, but they were (moral) authorities of the village."[9]

The above discussion suggests that the population rank order of a clan is a good proxy for the strength of informal institutions that a village leader can rely upon. Even if the sizes of clans are not drastically different, large clans (the largest clan in particular) are more likely to be well organized and enjoy greater social power. In the empirical analysis, our key explanatory variables are dummy variables indicating whether a VC or VPS came from the largest or second largest clan, which we believe summarize most of the information relevant to our study.[10]

**Village self-government, elections, and public investment.** Village self-government was reorganized by the CCP in the late 1970s after the abolition of the rural commune

[8]For clan members, social sanctions can take the form of a break of relationships, contempt, gossip, or even removal from the clan's family tree. For outsiders, non-compliance with the decision of powerful clans may also lead to unequal treatments in situations involving collective distribution.

[9]From the authors' interview in 2012.

[10]The size of lineage groups may also matter. In Appendix Section , we show that (1) our results are robust when we control for the VC's clan size and (2) the effect of informal institutions, as we measure them, varies little across clans with different sizes.

system. Village committees are designated as a "self-government organization" according to the Chinese Constitution. A village has two self-governing bodies: a village committee, which usually consists of three to seven members, and a village party branch, which includes several CCP members in the village. Village leaders are predominantly male. The VC, who has been democratically elected after village elections were introduced in the mid-1980s, leads the village committee. The position is also sometimes called the village chief or village head. The VPS leads the village party branch. Very often the village committee and village party branch overlap. Existing English and Chinese literature suggests that village officials are "sandwiched" between villagers and the township government, the lowest level of government (O'Brien and Li 1999, Oi and Rozelle 2000, Zhang 2007). They are supposed to be accountable to villagers, but they are also expected to fulfill tasks assigned by the township government.

Village elections first took place in Yishan County in Guangxi province as the People's Commune was dismantled in the early 1980s (Tan 2006: 59–63). Inspired by villagers' self-initiated acts, the CCP promoted village elections as an effort to address the information problem of holding local officials accountable and to improve local governance. To minimize risks, such as the state losing control of villages and compromising unpopular government policies, the government's democratization reform was gradual and highly controlled (O'Brien and Li 1999; Unger 2002). In 1987, a temporary version of the *Organizational Law of the Village Committee* (OLVC) was put into effect and village elections began to be formally introduced in most provinces. The formal version of the law was announced in 1998. Since then virtually all the villages have begun elections.

VCs are elected for three-year terms without term limits. Usually a handful of candidates are nominated in each election and a primary is held to reduce the number of candidates to two. The final round is run between these two front runners. Overt campaigning is not common in village elections (Pastor and Tan 2000; O'Brien and Han 2009). When elections were first introduced to villages, the township government maintained control of the nomination process. Only after 1998 when the OLVC was formally adopted were nominations open to all villagers.[11] The timing of the introduction of elections was largely determined by the provincial government's preferences.

One of the main jobs of the village committee is to provide village public goods (Whiting 1996; Oi and Rozelle 2000). It is responsible for determining public goods investment, as well as raising most of the funds required for the investment. Because the village committee does not have the legal authority to tax people, the only way it can finance public investment is through collecting fees and levies (hereafter, levies, for simplicity). Although levies were allowed by the central government before 2006, their amounts were usually small. Village leaders had to turn to villagers to ask for more levies if the village was to undertake a large public project. Unlike in more institutionalized contexts in which paying local taxes is enforced by law, village leaders had to exert a large amount of effort to convince villagers to pay levies. The collective action problem arises when villagers' semi-voluntary compliance is required for local public goods provision. This problem partly explains why scholars find that public goods were severely under-provided in rural China (e.g., Zhang et al. 2004; Luo et al. 2007, 2010).

## Data and Research Design

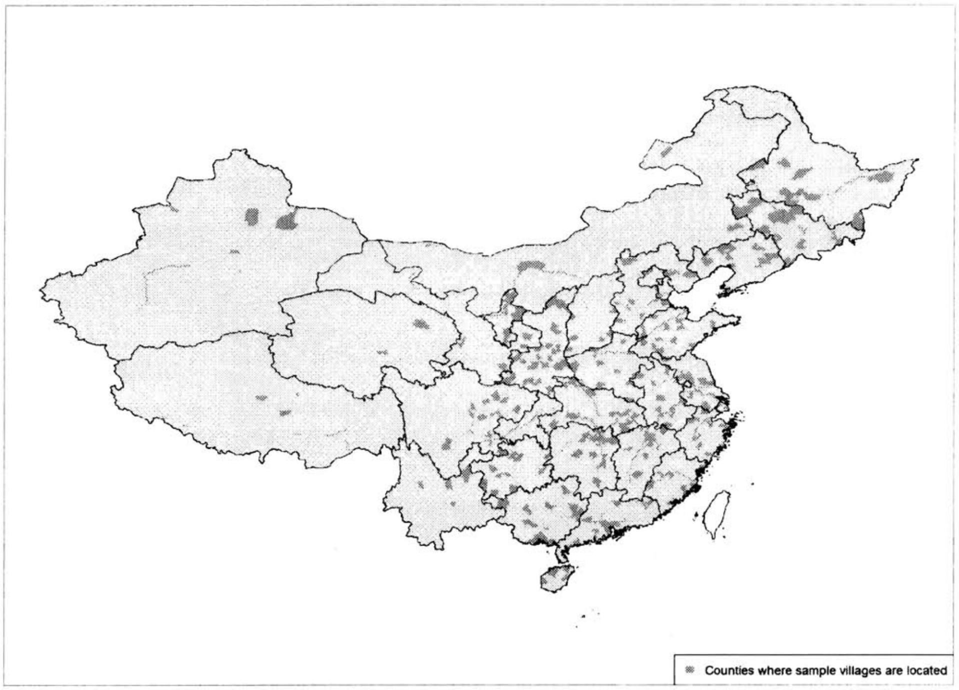
**Data.** This paper mainly uses a panel dataset of 220 villages from 1986 to 2005 from the Village Democracy Survey (VDS), a unique retrospective survey conducted by Gerard Padró i Miquel, Nancy Qian, and Yang Yao. The villages were selected from the sample of the National Fixed-Point Survey (NFS), a longitudinal survey maintained by China's Ministry of Agriculture.[12] We depict the locations of the sample villages (the counties that

[11]Nominations open to all villagers are popularly known as *haixuan* in China. It was first adopted in Lishu, Jilin in 1986 (Tan 2009).

[12]The NFS was started in 1986 to survey the same sample of households and villages over time. Except for 1992 and 1994, it provides annual data aggregated from daily household diaries. The NFS sample was first selected in 1986 according to a stratified random sampling strategy. Sample counties were first randomly selected from a province with the number of counties being proportional to the province's

they belong to) in Figure 1. In 2006, the VDS recorded the history of electoral reforms, traditional organizations, and public goods expenditure. In 2011, the VDS team returned to the same villages to collect data on village clan structure and more information on traditional organizations and elected village leaders.

Figure 1. SAMPLE VILLAGES




***Source:*** *The National Geomatics Center of China and the Village Democracy Survey.*

Data of electoral outcomes and public goods expenditure are obtained from village records, hence, concerns from report errors were minimal. Because the VDS only collected information of elected VCs, we only use observations in the post-election period to study the effect of informal institutions. However, focusing on the post-election period gives us the advantage of isolating the effect of informal institutions from that of electoral reforms.[13]

rural population. Then within a sample county, one village was randomly selected. Over the years, some villages dropped out of the survey mainly because they were incorporated into a nearby city, in which case a village in the same province was randomly selected to replace the dropped village. There are about 300 villages in the NFS. Among them, more than 220 villages have been in the sample for the 20-year period covered by this study. The VDS surveyed these villages. Martinez-Bravo et al. (2011) show that the VDS sample and the entire NFS sample are similar for a broad range of attributes.

[13]In Appendix Table 8, we perform a robustness check using data after 1995 to show that the timing

Information of lineage groups, including the identities of clans (surnames), their relative sizes as measured by shares in the village population, facilities they maintained, and activities and ceremonies they held, draw up on the collective responses of current and former living village leaders and elders, who were invited together to respond to the surveyors. The VDS recorded information on the four largest clans. Although there could be measurement errors in the exact number of villagers in each clan, villagers typically had consensus on the rank order of clan size in their villages. Therefore, we believe that the rank order was precisely recorded.[14] Moreover, because the Chinese government strictly prohibits permanent migration from rural areas, radical changes of the village social structure are less of a concern.[15] Since the VDS also recorded information of the elected VCs, we can identify the clan each VC belongs to by matching his surname with that of the clan's.[16] The VDS asked if a large clan kept records of its family trees (genealogies) or maintained a lineage hall. We use this information to form our measure for clan cohesiveness.

The VDS data are supplemented by annual data collected by the NFS. The control variables we use in this paper, such as village population, village household income, and village assets, come from the NFS. Data of levies that households paid to the village committee also come from the NFS.

The data we use have several merits. First, the information contains the most comprehensive data on village-level reform and governance outcomes in China. They cover a

of the electoral reform does not induce significant biases for the informal institution estimates. In 1995, most of the villages in our sample had adopted elections.

[14]In the survey, a meeting of the village leaders and elderly was convened in each village. Usually a handful of them came to meeting. The meeting lasted for about an hour, but consensus was very often quickly reached on the population rank order of the four largest clans. More time was spent on collecting information on the exact size of each of the four largest clans.

[15]Rural to urban migration soared at the beginning of the twenty-first century. We control for this factor in the regression analysis.

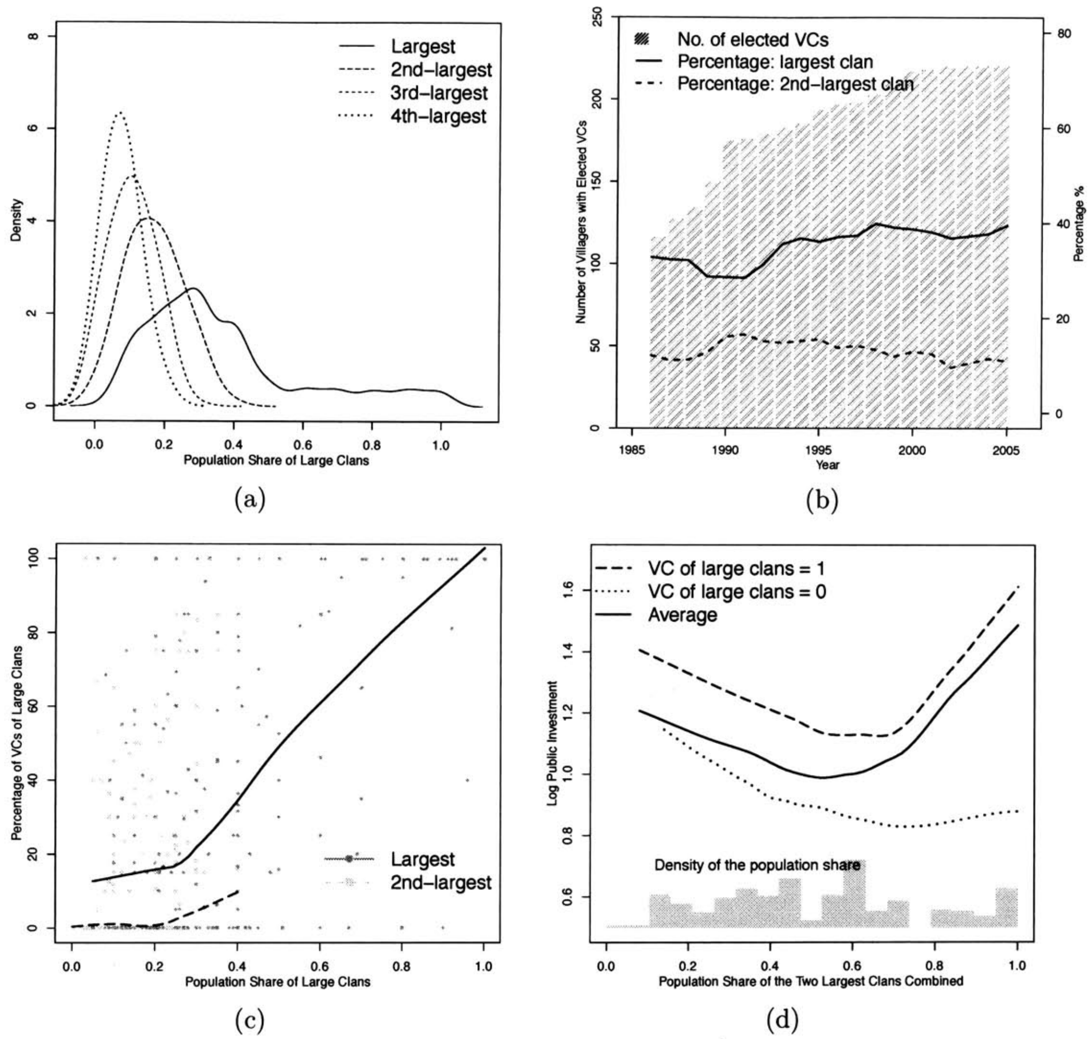
[16]In villages with only one surname (which are mostly in the south), we treat houses (*fang*) as separate lineage groups. Family names of women VCs did not reflect the clans they belonged to, because, in most cases, Chinese women do not change their family names after getting married. In the VDS, there were only 10 women elected as VCs in more than 1,000 recorded elections. We code them as coming from small clans. Dropping these observations does not affect our main results.

large and nationally representative sample and span a long period of time. Second, the panel structure, as well as the relatively large sample size, allows us to control for not only village and year fixed effects, but also time trends at the provincial level and even the village level. Village and year fixed effects account for unobserved time-invariant factors within each village and shocks that affect all villages in a given year, respectively. Time trends at the provincial or village level capture growing social and economic divergences across regions. Controlling for these factors eliminates a large number of potential confounders for the identification of the effect of informal institutions. For example, because village fixed effects allow us to make the comparison within villages, confounding factors associated with geography are effectively controlled for. Third, the quantitative data we have are mostly based on administrative records and, therefore, are comparable across villages and not likely to suffer from recall biases. Moreover, because the electoral outcomes and public investment data were collected directly from village records, they are not likely to be manipulated by village officials.

Figure 2a plots the distributions of the population shares of the four largest clans in sample villages. The average population shares of the largest and second-largest clans were 36 percent and 15 percent, respectively. In 2005, the average village in our sample had around 1,500 permanent residents. The average size of the largest clan in a village was thus around 400 villagers, or 100 households. Also, 81.8 percent of the villages did not have a lineage group that constituted the majority of the village population, and 74.5 percent of the villages had more than 10 surnames on the paternal side.

Figure 2b shows the onset of village elections since 1986. More than half of the villages adopted elections in 1986 and most villages had at least one election by the mid-1990s. The solid and dashed lines in this figure show the proportions of elected VCs coming from the village's largest and second-largest clans, respectively. On average, 35 percent and 13 percent of the VCs came from these two largest clans. Both numbers remained relatively stable over time. Even though lineage groups might have a big impact on local

Figure 2. Large Clans, Elections, and Public Investment




***Note:*** *Figure 2a shows the distributions of the population share of the four largest clans in each village. Figure 2b shows the roll-out of village elections (left axis) and the percentages of elected VCs coming from the largest and second-largest clans (right axis). Figure 2c shows the percentages of VCs from the largest and second-largest clans over their respective (relative) clan size with two loess fits. Figure 2d shows the relationship in raw data between VCs of large clans and the amount of public investment. The x-axis is the combined population share of the largest and second-largest clans in a village.*

governance, they did not necessarily dominate village elections. Figure 2c plots the share of VCs of the two largest clans against their respective clan size. It is clear that large clans were not over-represented.[17] There are also large cross-sectional heterogeneities;

[17] Su et al. (2011) show that clan networks in rural China mobilize voters to go to voting stations, but there is not enough empirical evidence suggesting that large clans dominate village elections. Anecdotal

some villages elected VCs of large clans all the time while others never did. This occurs probably because in some places large clans are well organized, while in other places, the upper-level government has a big influence on putting its favored candidates on ballots or getting them elected.

Public investment falls into six categories: schooling, roads and sanitation (water supply and sewer systems), electric power, irrigation, forestation, and others. Figure 2d plots the average log public investment against the combined population share of the two largest clans. The plot exhibits a non-monotonic relationship between public goods expenditure and the size of the two largest lineage groups. There was more public goods expenditure in the most and least homogenous villages.[18] Such a relationship may be misleading, though, because the figure does not control for other variables. For example, many of the least homogeneous villages are located in coastal regions, so more investment in these villages could be because of higher levels of economic development. However, if we compare the amount of public investment during the terms of VCs from large clans and the amount during the terms of VCs from smaller clans, we find a clear gap between the two: when VCs of large clans were in office, there was more investment. Figure 2d illustrates that the gap was relatively stable across villages of different clan structures. Since we are looking at the raw data, though, this gap consists of variations both across and within villages.

As mentioned earlier, among the elected VCs, 35 percent and 13 percent were from the largest and second-largest clans, respectively. The average VC was around 42 years old when he was elected and had received 6.4 years of formal education. Three quarters of them were CCP members and 56 percent were already village cadres when they ran

evidence suggests that the CCP is constantly worried about the possibility that clans capture rural politics and has tried different measures to prevent it from happening (Mattingly 2014).

[18]The fact that more homogenous villages had a higher level of public goods expenditure is consistent with a wealth of literature on ethnic homogeneity and public goods provision, for example, Alesina, Baqir and Easterly (1999), Alesina, Baqir and Hoxby (2004), and Habyarimana et al. (2009), among many others.

Table 1. DESCRIPTIVE STATISTICS

| **Village-Year Observations** | Obs. | Mean | S.D. | Min. | Max. |
|---|---|---|---|---|---|
| Any public investment project | 3,742 | 0.23 | 0.42 | 0 | 1 |
| Log total public investment (1,000 yuan) | 3,742 | 1.09 | 2.15 | 0 | 10.60 |
| Log village-average household levies (yuan) | 1,080 | 4.22 | 1.90 | 0 | 7.06 |
| Log village population (persons) | 3,513 | 7.20 | 0.61 | 4.67 | 9.16 |
| Log net income per capita (yuan) | 3,513 | 7.22 | 0.83 | 1.86 | 10.42 |
| Log village asset (yuan) | 3,513 | 9.01 | 1.62 | 2.67 | 15.35 |
| Average household size (persons) | 3,513 | 3.93 | 0.59 | 2.00 | 6.39 |
| Arable land per capita (mu) | 3,513 | 1.75 | 1.88 | 0.004 | 16.20 |
| Log number of people migrating out of the village | 2,685 | 2.19 | 1.10 | 0.00 | 5.50 |
| Log taxes to the upper-level government (1,000 yuan) | 2,530 | 2.27 | 1.86 | 0.00 | 8.80 |
| Log transfers from the upper-level government (1,000 yuan) | 2,530 | 1.14 | 1.61 | 0.00 | 7.50 |
| Share of administrative expenditure in total expenditure | 3,037 | 0.23 | 0.22 | 0.00 | 1.00 |
| Contested election | 3,742 | 0.77 | 0.42 | 0 | 1 |
| Open nomination | 3,742 | 0.70 | 0.46 | 0 | 1 |
| Secret ballot | 3,742 | 0.38 | 0.49 | 0 | 1 |
| Proxy voting | 3,742 | 0.71 | 0.45 | 0 | 1 |
| Moving ballot boxes | 3,742 | 0.68 | 0.47 | 0 | 1 |
| **Village Chairpersons (by Term)** | Obs. | Mean | S.D. | Min. | Max. |
| VCs from the largest clan | 1,315 | 0.36 | 0.48 | 0 | 1 |
| ~ from the second-largest clan | 1,315 | 0.13 | 0.33 | 0 | 1 |
| ~ of large clans (from either the first or the second) | 1,315 | 0.48 | 0.50 | 0 | 1 |
| Years of education | 1,210 | 6.39 | 2.30 | 0 | 13 |
| Age when running election | 1,203 | 41.56 | 8.72 | 19 | 90 |
| CCP member | 1,195 | 0.75 | 0.43 | 0 | 1 |
| Village cadre before election | 1,209 | 0.56 | 0.50 | 0 | 1 |
| Managerial jobs before election | 1,209 | 0.02 | 0.14 | 0 | 1 |
| Experience of running election | 1,205 | 0.71 | 0.46 | 0 | 1 |
| Family background: poor peasant | 1,213 | 0.79 | 0.41 | 0 | 1 |
| Denounced in Culture Revolution (*pidou*) | 1,203 | 0.05 | 0.22 | 0 | 1 |
| Relative vote share of VCs of large clans | 795 | 0.51 | 0.44 | 0.00 | 1.00 |
| VPS of large clans | 870 | 0,51 | 0.50 | 0 | 1 |
| VC and VPS in the same clan | 1,315 | 0.24 | 0.42 | 0 | 1 |
| Serving as VPSs ("one-shoulder") | 1,315 | 0.08 | 0.27 | 0 | 1 |
| In the village party branch | 830 | 0.62 | 0.49 | 0 | 1 |
| **Sample Villages** | Obs. | Mean | S.D. | Min. | Max. |
| No. of clans (surnames) | 220 | 26.76 | 23.44 | 1 | 150 |
| Population share of the largest clan | 220 | 0.36 | 0.23 | 0.05 | 1 |
| Population share of the second-largest clan | 220 | 0.17 | 0.08 | 0.00 | 0.40 |
| Population share of the third-largest clan | 220 | 0.10 | 0.07 | 0.00 | 0.30 |
| Population share of the fourth-largest clan | 220 | 0.07 | 0.05 | 0.00 | 0.20 |
| Records of family trees (of the two largest clans) | 200 | 0.48 | 0.50 | 0 | 1 |
| Lineage halls (of the two largest clans) | 200 | 0.17 | 0.37 | 0 | 1 |

for office. Among the 200 sample villages that have detailed information of large clans, 48 percent had a large clan keeping records of family trees; 17 percent had a large clan maintaining a lineage hall. Table 1 presents descriptive statistics of the variables used in the regression analysis, including the number of observations, mean, standard deviation, minimum, and maximum of each variable.

Our main explanatory variable is constructed on the population rank order of clans because we believe that large clans have richer and stronger informal institutions. Here we present some supporting descriptive evidence. Figure 3 illustrates the level of organization for the four largest clans in the sample villages. On average, the largest and second-largest clans were more likely to have maintained lineage halls and to hold clan ceremonies on a yearly basis. For example, among all largest and second-largest clans in the sample villages, 15.5 percent and 11.1 percent had lineage halls, respectively. In comparison, the numbers for the third- and fourth-largest clans are around 7.5 percent. Moreover, the probability of having a lineage hall that was established before the reform era is much higher for the largest clan than for other clans. In our empirical analysis, we focus on the largest and second largest clans to allow sufficient modelling flexibility.[19]

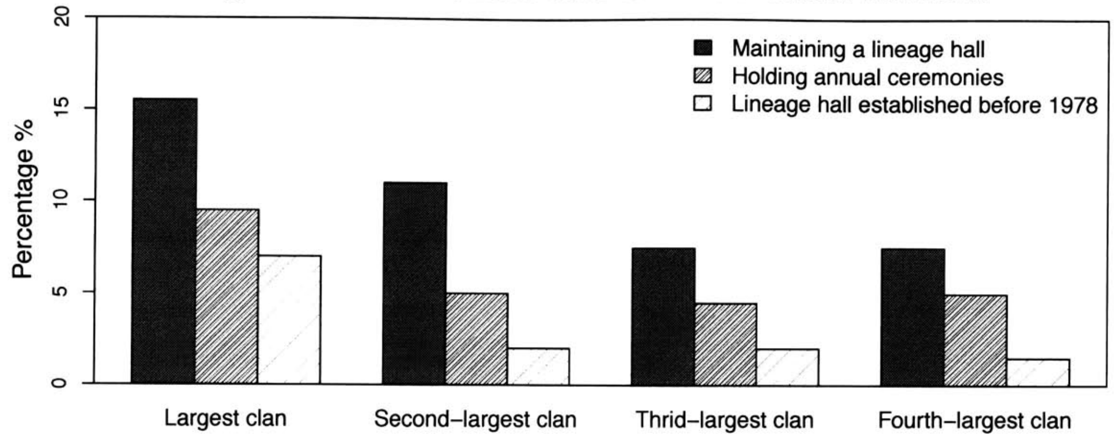
**Main identification strategy.** Our key independent variables are binary indicators of whether an elected VC came from the largest or second-largest clan in the village. Our baseline specification is the following fixed effects model:

$$y_{it} = \beta_1 D_{it,1} + \beta_2 D_{it,2} + \eta_i + \delta_t + \epsilon_{it}, \tag{1}$$

where $y_{it}$ is the outcome variable (e.g., the log amount of public investment) for village $i$ in year $t$; $D_{it,1}$ and $D_{it,2}$ are dummy variables indicating whether a VC was from village $i$'s largest or second-largest clan in year $t$, respectively; $\eta_i$ and $\delta_t$ are village and year fixed

[19]Appendix Table 10 shows that our main findings remain unchanged if we define the key independent variable solely based on the largest clan.

Figure 3. Large Clans and Lineage Organizations




***Note:*** *This figure shows the (1) percentage of having a lineage hall, (2) percentage of holding annual ceremonies in the past five years, and (3) percentage of having a lineage hall established before 1978, of the four largest clans in the sample villages.*

effects; and $\epsilon_{it}$ represents idiosyncratic shocks. The village and year fixed effects absorb time-invariant heterogeneities across villages and aggregate shocks that affect all villages in a given year, respectively. The identifying assumption is that $D_{it,1}$ and $D_{it,2}$ are not correlated with the error terms $\{\epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{iT}\}$. In other words, we assume that the choice of VCs is quasi-random with respect to the amount of public goods expenditure after both the independent and dependent variables are demeaned within each village and across villages in a given year. The parameters we are concerned about are $\beta_1$ and $\beta_2$; we expect that they are significantly positive.

We also add four sets of additional controls to the baseline specification. First, we control for provincial linear time trends to capture regional economic divergence. Second, we replace these trends with village-specific linear time trends to take into account trending factors at the village level. Third, we control for time-varying covariates from NFS to show that our finding is not driven by these variables. The covariates will be introduced later when we present the relevant robustness results. Fourth, we control for taxes/fees the village committee paid to the upper-level government and total transfers it received from it to capture the influence of the upper-level government. We also control for the number of villagers migrating out each year. However, these specifications may not rule out the impacts of other unobserved time-varying variables that are correlated with the choice of VCs and public goods expenditure at the same time. We will address this concern later using a regression discontinuity design.

When conducting robustness checks and exploring mechanisms, we also use the following simplified specification:

$$y_{it} = \beta D_{it} + \eta_i + \delta_t + \epsilon_{it}, \tag{2}$$

where $D_{it}$ is a dummy variable indicating whether a VC was from village $i$'s largest clan or second-largest clan in year $t$. As we will see from the baseline results, both $\beta_1$ and $\beta_2$ are indeed large and positive, and they are statistically indistinguishable from each other in most cases. Using the above simplification, therefore, does not lose much information.

## Main Results

This section presents the baseline empirical results and some robustness checks. The main outcome of interest is public goods expenditure. We focus on the association between VCs of large clans and the amount of public investment during their terms in office. The dependent variable is the log amount of village investment (1,000 yuan).[20] Table 2 shows the baseline results, which are produced by the estimation of Equation 1 (except for Column 1). In Column 1, we show the raw result from an ordinary least squares (OLS) regression without controlling for village fixed effects; the estimated coefficients of both VC dummies are positive. In Column 2, when both year and village fixed effects are controlled for, the coefficients of the two VC dummies are 0.412 and 0.303, respectively. Both are statistically significant at the 5 percent level. This means that a VC from the two

[20]The dependent variable is generated by `log(x+1)`, in which `x` is the amount of public investment, because investment can be zero in a year.

largest clans is associated with 35 to 51 percent more expenditure in public investment. In Column 3, we control for provincial linear time trends; the estimates remain stable. In Column 4, provincial linear time trends are replaced by village-specific linear time trends. The estimates of interests are 0.359 and 0.256, similar to the baseline results. The standard errors go up quite a bit, and the dummy for the second-largest clan turns only marginally significant. In Column 5, we go back to provincial linear time trends, but add five time-varying control variables from the NFS, namely, log village population, average village household size, arable land per capita, log income per capita, and log assets owned by the village committee. These controls capture the size, demographics, agricultural endowment, and economic resources of the village. The results are very similar to those in Column 2.

Table 2. VCs of Large Clans and Village Public Investment

| | Log Public Investment (1,000 yuan) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | OLS | FE | FE | FE | FE | FE |
| VC of the largest clan | 0.332 | 0.412 | 0.379 | 0.359 | 0.378 | 0.481 |
| | (0.126) | (0.148) | (0.148) | (0.189) | (0.157) | (0.200) |
| VC of the second-largest clan | 0.183 | 0.303 | 0.328 | 0.256 | 0.367 | 0.421 |
| | (0.151) | (0.148) | (0.145) | (0.193) | (0.155) | (0.227) |
| Dependent variable mean | 1.092 | 1.092 | 1.092 | 1.092 | 1.083 | 1.225 |
| Year fixed effects | x | x | x | x | x | x |
| Village fixed effects | | x | x | x | x | x |
| Provincial linear trends | | | x | | x | x |
| Village linear trends | | | | x | | |
| NFS controls | | | | | x | x |
| Persons migrating out | | | | | | x |
| Taxes to the upper-level government | | | | | | x |
| Transfers from the upper-level government | | | | | | x |
| Observations | 3,742 | 3,742 | 3,742 | 3,742 | 3,513 | 2,530 |
| Villages | 220 | 220 | 220 | 220 | 217 | 208 |

***Note:*** *This table shows that the presence of a VC of large clans is associated with a larger amount of village public investment. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) during that year. The independent variables are two dummy variables indicating whether a VC came from the village's largest or second-largest clan, respectively. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced.*

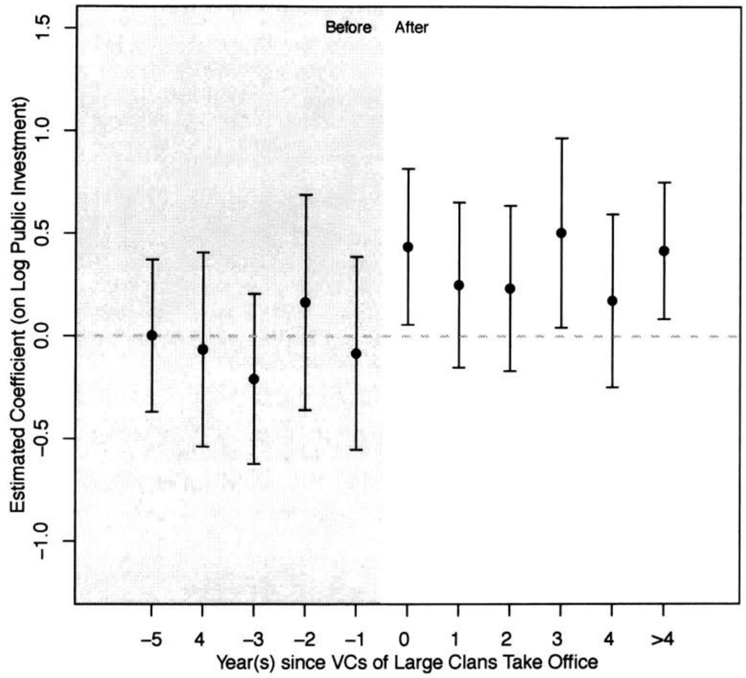
Next, we consider the relationship between the village committee and the upper-level government. Two possibilities might affect the village committee's ability to provide public goods. First, VCs of large clans might have better access to government funds, which were often crucial for investment projects. Second, because the village committee was obligated to follow directives coming from the township government, the amount of money the village committee paid to the township might have a great impact on the village committee's budget constraints. Because of these concerns, in Column 6, we additionally control for log total transfers the village committee received from the upper-level government and log total taxes and fees it handed over to the upper-level government each year. Moreover, to account for the impact of high waves of rural to urban migration since the beginning of the twenty-first century, we also add the total number of people migrating out of the village each year in the regression. All three variables are available for 208 villages after 1993. We find that the coefficients of VC of large clans become even bigger.[21]

In summary, the estimated coefficient of VC of the largest clan is very robust, remaining significant and varying only slightly when different controls are added. The coefficient of VC of the second-largest clan is also robust unless village-specific linear time trends are controlled for. These results show that the association between VCs of large clans and public goods expenditure is robust and not likely to be driven by trending factors, village-level economic and demographic changes, or differentiated support from the upper-level government. Because the coefficients of VC of the largest clan and VC of the second-largest clan are statistically indistinguishable from each other in most cases, in the rest of the paper we use the simplified model of Equation 2.

To establish a causal relationship between VCs of large clans and public goods expenditure, we need to be sure that the identifying assumption is valid. We are more confident that this assumption holds if we find that public goods expenditure increases

[21]The county government started taking charge of village public goods provision after the agricultural taxes, as well as village levies, were formally abolished in 2006. In Appendix Tables 8 and 9, we conduct more robustness checks to show that our finding is robust in different time periods and is not driven by extreme values.

Figure 4. DYNAMIC EFFECT OF VCS OF LARGE CLANS ON PUBLIC INVESTMENT




***Note:*** *This figure shows the dynamic effect of VC of large clans on the amount of public investment. Each black dot is an estimated coefficient of a dummy variable indicating the year(s) since the most recent VC of large clans took office (or before he took office).*

right after VCs of large clans took office. To achieve this, we create a set of dummies $d_k$, $k = -5, -4, \cdots, 0, 1, \cdots, 4$ where $k = 0$ indicates the year a VC of the two largest clans took office, and other values respectively correspond to a specific year relative to that year. For example, $k = -1$ indicates one year before the closest year when a VC of a large clan replaced a VC of a small clan, and $k = 1$ indicates one year after. All the years that were five or more years before are pooled together as the reference category indicated by $k = -5$, while $k = 4$ includes four or more years after the year a VC of a large clan took office. Then we estimate Equation 2 by substituting this new set of dummies for $D_{it}$. The estimated coefficients of the dummies are shown in Figure 4. Before VCs of the two largest clans took office, the estimates are mostly negative and statistically insignificant. The coefficients turn positive and statistically significant only after VCs of large clans took office.[22]

Because the VDS recorded the amount of investment by project type, we can check for which types of investment the association between VCs of large clans and public goods expenditure is stronger. The results based on Equation 2 are shown in Table 3. The dependent variable is the log amount of village investment by type. Table 3 suggests that strong associations exist between VCs of large clans and investment in facilities of village primary schools and irrigation infrastructure. Although irrigation infrastructure can be built only for the benefits of large clans, village primary schools are rarely discriminatory in rural China. Therefore, we can at least conclude that having village leaders from large clans also benefited the rest of the villagers in addition to clan members. Moreover, in the long run, we do not observe that the level of income inequality deteriorated more quickly in villages with large lineage groups. The problem of clan capture seems to be less severe than one would otherwise expect.

**Clan cohesiveness.** To provide further evidence that it is the informal institutions of large clans that matter, we examine additional information on large clans in the sample villages. When a clan is more cohesive, it is more likely that it has greater social power in the village, as a result, its rules are more strictly enforced. Therefore, we expect the association between VCs of large clans and public goods expenditure to be stronger in villages with more cohesive large clans.

To test this hypothesis, we look at two indicators of clan cohesiveness: (1) whether the largest or second-largest clan kept records of family trees, and (2) whether they maintained lineage halls. We take these two variables as proxies for clan cohesiveness because they signify how closely clan members were connected with each other and whether a clan had sufficient organizational capacity. Records of family trees and lineage halls are

[22] Note that the coefficient is still positive and significant three years after a VC of the largest two clans took office although a VC's term is three years. One possibility is that VCs of the two largest clans stayed in office for more than one term. Another possibility is that the successor also came from the two largest clans.

Table 3. VCs of Large Clans and Village Public Investment: by Project Type

| | Log Public Investment (1,000 yuan) | | | | | |
|---|---|---|---|---|---|---|
| | Schooling | Road & Sanitation | Electricity | Irrigation | Foresta-tion | Others |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | FE | FE | FE | FE | FE | FE |
| VC of large clans | 0.161 | 0.061 | 0.070 | 0.148 | 0.014 | 0.057 |
| | (0.061) | (0.066) | (0.041) | (0.054) | (0.030) | (0.055) |
| Dependent variable mean | 0.292 | 0.358 | 0.185 | 0.211 | 0.050 | 0.176 |
| Year and village fixed effects | x | x | x | x | x | x |
| Observations | 3,742 | 3,742 | 3,742 | 3,742 | 3,742 | 3,742 |
| Villages | 220 | 220 | 220 | 220 | 220 | 220 |

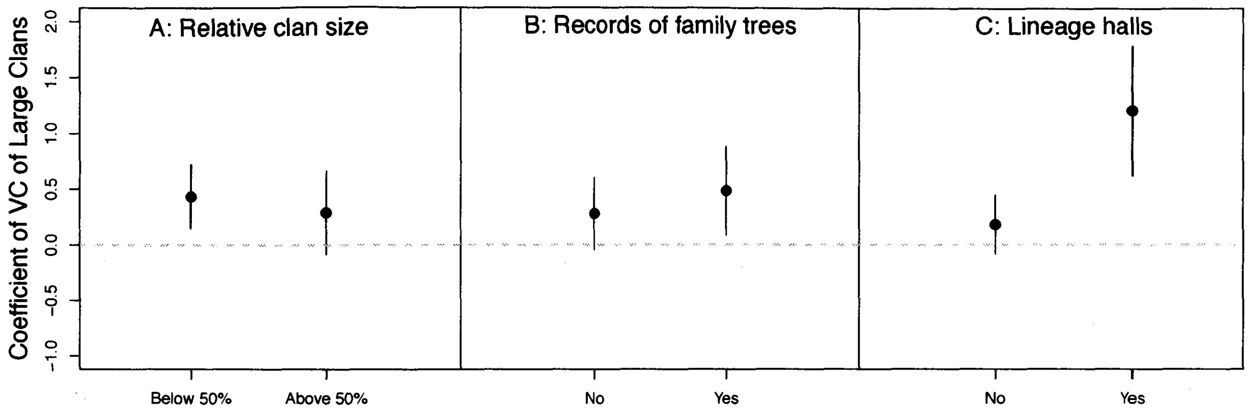
***Note:*** *This table shows the associations between a VC of large clans and village public investment by project type. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) of each type during that year. The independent variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clans. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

specific to clans and signal a close relationship within the clan and frequent clan activities. Annual sacrificing activities, weddings, funerals and other clan events often take place in lineage halls.

It is possible that in villages with more public investment and better infrastructure, lineage halls could be more regularly refurbished for reasons that we cannot fully control for. To minimize such biases, the indicator of lineage halls is coded as one if they were built before the observed time periods and zero otherwise. Because maintaining records of the family tree requires persistent efforts of clan members, it is less likely to incur such biases.

We interact VC of large clans with each of the two indicators of clan cohesiveness and put both the VC dummy and the interaction term in regressions using the baseline fixed effects specification. As a comparison, we also use a specification that includes the interaction between the VC dummy and a dummy variable indicating that the combined size of the two largest clans was above 50 percent (roughly the median). Figure 5 visualizes

Figure 5. VCs of Large Clans, Clan Size, and Cohesiveness




***Note:*** *The above figures show the heterogeneous effects of VC of large clans on public goods expenditure with 95% confidence intervals. From left to right, the sub-samples are (1) villages whose combined size of the two largest clans is above or below 50 percent, (2) villages in which any of the two largest clans had kept records of family trees or not, and (3) villages in which the two largest clans had maintained any lineage halls since the beginning of the observed time periods or not.*

the results. Panel A of Figure 5 shows that the association between public goods expenditure and VCs of large clans is not increasing in the combined size of the two largest clans. However, Panels B and C of Figure 5 show that when large clans appeared to be more cohesive, i.e. having maintained records of family trees and especially lineage halls (in 48% and 17% of the villages, respectively), VCs of large clans are strongly associated with more spending on public investment. These results suggest that what really matters for spending on public investment is not the number of people a large clan had, but the social aspect of the organization, likely the rules and norms it enforced.

**The role of village party organizations.** Existing literature shows that VCs can face considerable constraints when exercising power (e.g., Oi and Rozelle 2000, O'Brien and Han 2009). A VC not only receives orders from the township government, but is also subject to checks and even directives of the village party organization, especially the VPS. In fact, studies show that the power struggle between the VC and the VPS paralyses village self-government in some places (e.g., Tan 2010). Would the consideration of VPSs alter our main findings? For instance, would the VPS's clan membership affect the VC's ability to provide public goods? What would happen if the VC was also the VPS, which is called *yijiantiao* (literally, "one shoulder")? What if the VC and VPS were from the same clan? Or what if the VC was in the village party branch, a sign that he was recognized and supported by the VPS?

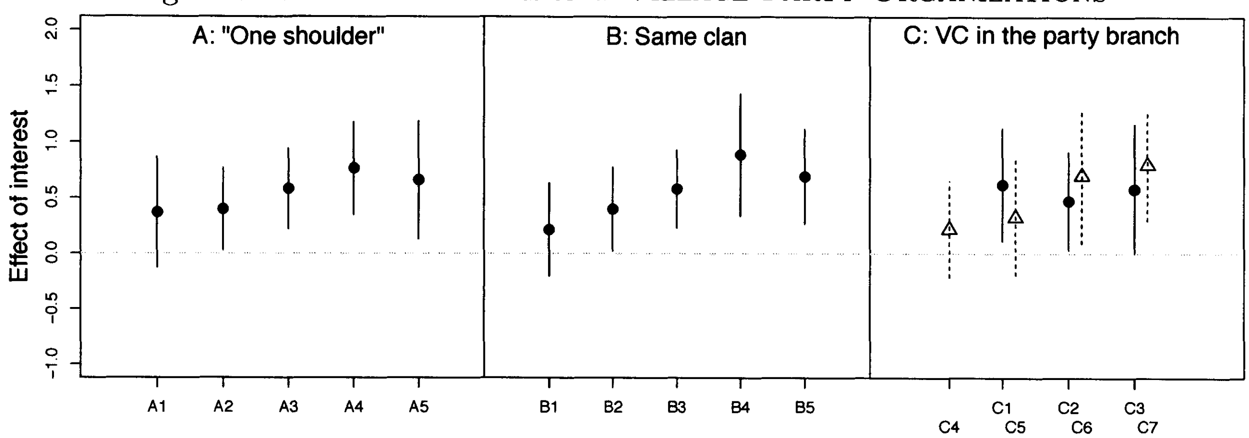
Fortunately, the VDS includes data on the VPS and village party organizations for more than 130 villages, roughly 60% of the entire sample.[23] Such information allows us to answer the questions we just posed. Using names of VPSs and data on village clan structure, we define a dummy variable indicating whether the VPS came from the village's largest or second-largest clan.

We first consider how "one shoulder" affects our results. For that purpose, we define a dummy variable indicating "one shoulder." We include the VC dummy, the VPS dummy, the "one shoulder" dummy, as well as the interactions between the VC and VPS dummies and between the VC and "one shoulder" dummies in the baseline two-way fixed-effect model and visualize the result in Panel A of Figure 6. Our specification allows us to compare five scenarios with the reference scenario in which both the VC and the VPS came from small clans and were not the same person: (A1) the VC and VPS were the same person but he was not from one of the two largest clans, (A2) the VPS came from one of the two largest clans while the VC came from a small clan, (A3) the VC came from one of the two largest clans while the VPS came from a small clan, (A4) both the VC and the VPS were from one of the two largest clans yet they were not the same person, and (A5) the VC and VPS were the same person and came from one of the two largest clans. Figure 6 shows that the average amounts of public investment of the last four scenarios are significantly higher than that of the reference scenario, after village and year fixed effects have been controlled for. The effect under the first scenario, i.e., "one shoulder"

[23]Lack of data on VPSs and party organizations for the rest of the VDS sample was due to administrative reasons. Statistical analysis shows that villages with available data are not substantially different from the rest.

from a small clan, is positive but not statistically significant. Moreover, among the five scenarios, the level of public goods expenditure is highest under the fourth scenario, in which the VC and the VPS, though not the same person, were both from large clans. Those results indicate that "one shoulder" is less important than the clan membership of the VC and the VPS.

Figure 6. VILLAGE LEADERS AND VILLAGE PARTY ORGANIZATIONS




***Note:*** *This figure shows the heterogeneous effects of VCs of large clans on public goods expenditure with 95% confidence intervals. From left to right, we consider three cases: (1) whether the VC and VPS were the same person, (2) whether the VC and VPS came from the same clan; and (3) whether the VC was in the village party branch.*

Next, we consider the effects when the VC and the VPS came from the same clan. The procedure is similar. Again, we estimate a "fully saturated" model using the baseline two-way fixed-effect specification and present the result in Panel B of Figure 6. The reference scenario, in which the VC and VPS were from small and distinct clans, is compared with the following five scenarios: (B1) the VC and VPS were from the same small clan; (B2) the VPS was from one of the two largest clans while the VC was not; (B3) the VC was from one of the two largest clans while the VPS was not; (B4) the VC and VPS were from large yet distinct clans; and (B5) both the VC and VPS were from the same large clan. The last four comparisons give positive and significant estimates but the first does not. The highest level of public goods expenditure happened when the VC and the VPS came from distinct large clans (Scenario B4). That is, the VC and the VPS did not have to come from the same clan as long as both of them came from a large clan. [24]

Lastly, we investigate whether being in the village party branch enhances a VC's ability to provide public goods. Again, we use a "fully saturated" model controlling for village and year fixed effects. Panel C of Figure 6 shows the results. The reference scenario is that the VC was not in the village party branch and neither the VC nor the VPS was from one of the two largest clans. There are seven scenarios to be compared with. In the first three scenarios, we have the VPS not from one of the two largest clans (the estimated effects of interest are depicted with dots) with one of the following three cases: (C1) the VC was the in the village party branch; (C2) the VC was from one of the two largest clans; and (C3) the VC was from one of the two largest clans and in the village party branch. The other four scenarios, whose estimated effects of interest are depicted with triangles, are when the VPS was from one of the two largest clans with one of the following four cases: (C4) the VC was neither from a large clan nor in the village party branch; (C5) the VC was in the village party branch while not from one of two largest clans; (C6) the VC was from one of the two largest clans while not in the village party branch; and (C7) the VC was from one of the two largest clans and in the village party branch. The estimated effects of interest are positive and statistically different from zero in all but Scenarios C1 and C5, in which the VC was in the party branch but not from one of the two largest clans. Also worth noting are that (1) all scenarios in which the VC was from one of the two largest clans have significantly positive estimates; (2) it is not necessary to require the VPS coming from one of two largest clans to have more public goods expenditure as long as the VC was from one of the largest clans, a clear result when Scenario C1 is compared with Scenario C5; and (3) the effect of VPS coming from one of the two largest clans becomes insignificant if the VC did not, a result shown by

[24]The above two sets of results suggest that it seems a good thing if there existed some competition between the VC and the VPS as long as they came from one of the two largest clans. Further exploration is needed to find out the exact reason behind it.

Scenario C4.[25]

In summary, we not only show that the strong association between VCs of large clans and a higher level of public investment is robust when we take into account the roles of VPSs and village party organizations, but also find that the level of public investment is higher when the VPS was from a large clan than when he was not. However, we do not find enough evidence that the VC and VPS being the same person or from the same clan brought about additionally more public investment. Nor do we find that the VC being in the village party branch is particularly important for public goods provision once we control for clan memberships of the VC and VPS. Lastly, the role of the VPS diminished when the VC came from a large clan. In the rest of the paper, we will mainly focus on the role of informal institutions associated with VCs primarily because: (1) we have more complete data on VCs than VPSs, and (2) there are potentially more quasi-exogenous variations in the turnovers of VCs than VPSs—as we will see in the next section, these variations give us more leverage to identify the causal effect of informal institutions of lineage groups.

## A Regression Discontinuity Design

In this section, we employ a regression discontinuity (RD) design to address the potential endogeneity of electoral outcomes. Recall that we rely on elections as the source of variations of informal institutions that affect local governance. A natural question is why sometimes the largest clans won the election, while at other times they lost. We admit that the impact factors are complex and mostly beyond our knowledge. One obviously important factor is the CCP. To strengthen its rule in the countryside, the CCP has been trying to demobilize clans in elections. Our key identification assumption, though, is that those factors are uncorrelated with public investment, the outcome variable. Our fixed

[25]These results arise probably because VCs, who were popularly elected, were more able to obtain support from their clans than VPSs, who were appointed by the government.

effects approach and additional controls of the provincial and village time trends, as well as other time-varying covariates, buttress this assumption. However, concerns of reverse causality and unobserved time-varying confounders remain. For example, villagers may expect VCs of large clans to provide more public goods and, therefore, elect them into office. A sharp RD design can address this concern because after conditioning on the forcing variable, the treatment indicator is uncorrelated with time-varying confounders at the cutoff. In our case, the forcing variable is the share of votes of a candidate from the largest or second-largest clan against the share of votes of a contender from a smaller clan.

Several caveats of employing an RD design in this study are worth noting. First, an RD estimate gives the local average treatment effect at the cutoff, which is 50 percent of all the votes. This quantity is not necessarily a quantity of interest and can provide very different estimates from those generated by the benchmark fixed effects models. One might be especially concerned about the external validity of an RD design in the Chinese context. Because formal democratic institutions are weak, elections may not be "allowed" to be close under many circumstances.

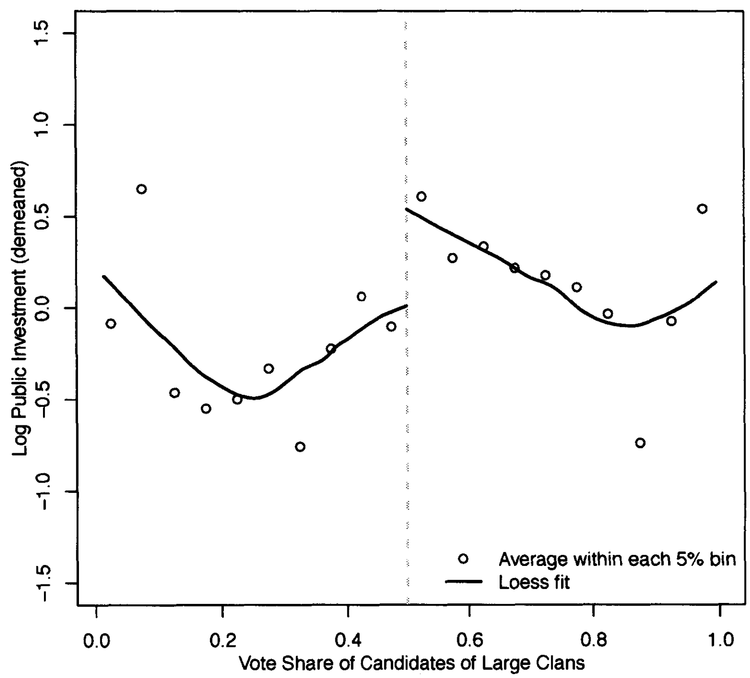
Second, although the treatment assignment mechanism is very clear in an RD design, in reality, the assignment mechanism may suffer from manipulation of the forcing variable by interested parties, making the RD design invalid. Evidence exists that an RD design fails regarding elections of US House of Representatives during a certain period of time (Caughey and Sekhon 2011).[26] Vote-buying, electoral frauds, and interference of the upper-level governments were widely observed by scholars for Chinese rural elections (e.g., Shi 1999). Therefore, we may need to worry about the validity of an RD design in the context of rural China.

Third, an RD design typically demands a large amount of data. To construct an RD

[26]A follow-up study shows that the problem is not as severe as one might think and RD remains a valid method for causal inference in most situations (Eggers et al. (forthcoming)).

design, we need not only information of the elected VCs, but also information of their runoffs, including the lineage groups the latter belonged to and the votes they received in elections. These requirements cut the sample size to 2,230 village-year observations and 871 elected terms, compared with the original 3,742 observations and 1,315 terms. Dropped observations are mostly in early periods of the time series when village elections were not contestable (therefore no runoff information was recorded). Moreover, because we are interested in the effect of VCs of large clans and use VCs from small clans as comparison, only observations in which one of the candidates was from one of the two largest clans while the other was from a smaller clan are useful for constructing the RD design. This requirement further reduces the sample down to 715 observations and 253 terms.[27] Because the identification comes from close elections, the power of our RD analysis is limited.

Figure 7. ROBUSTNESS CHECK: A REGRESSION DISCONTINUITY DESIGN




***Note:*** *This Figure shows the averages of log amount of investment within each 5 percent vote-share bin and two loess fits from locally linear regressions on both sides of the cutoff.*

[27]To remove time-invariant heterogeneities and aggregate shocks, we first run a standard fixed effects model controlling for village and year fixed effects and use the residuals in the RD analysis.

Bearing these concerns and limitations in mind, we present the main result of the RD design in Figure 7, which shows the averages of log investment within each 5 percent vote-share bin and two loess fits (from locally linear regressions) on both sides of the cutoff. The RD estimate is 0.573 with a standard error of 0.301; both are almost twice as large as the fixed effects estimates.[28] We find that the results from the RD design are consistent with our main finding and offer us more confidence that VCs of large clans causally increased public goods expenditure.

## Mechanisms

We have already shown that the presence of VCs from one of the two largest clans is associated with at least 35 percent more investment in public goods. In this section, we investigate two mechanisms, namely, the collective action mechanism and the accountability mechanism, through which informal institutions could possibly facilitate public goods provision.

First, we test whether the presence of VCs of large clans is connected with easier collective action among the villagers by using household-level data of levies that villagers paid to the village committee. As mentioned before, a VC needed to seek villagers' voluntary compliance to collect levies from them. If VCs of large clans were more likely to collect more levies for public investment than VCs from small clans, we then have a critical piece of evidence to support the collective action mechanism.

We have household-level data for around one-third of the sample villages.[29] Table 4 presents the results based on this sample. Using the baseline model that controls for both village and year fixed effects, Column 1 shows that the presence of VCs of large clans is

[28]In Appendix Table 14 and Figure 11, we present the point estimates from the RD analysis and conduct more validity tests.

[29]The household-level data come from the NFS which only allows researchers to obtain a maximum of one third of its household data. In addition, it does not allow the household-level data to be transported and used directly outside China; the data were first processed in China to generate means and values at each income decile for the variables of interest. Our analysis is, therefore, based on the processed data.

Table 4. VCs of Large Clans and Levies

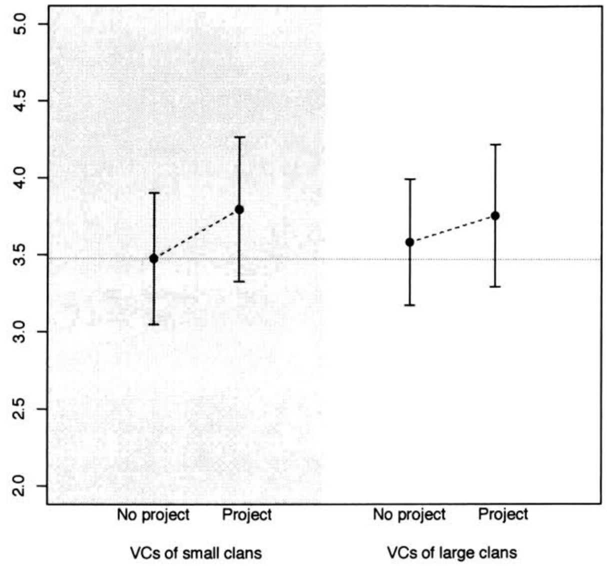
| | Log Levies (yuan) | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| | FE | FE | FE |
| VC of large clans | 0.132 | | 0.110 |
| | (0.192) | | (0.188) |
| Public investment dummy | | 0.304 | 0.321 |
| | | (0.096) | (0.134) |
| VC of large clans × public investment dummy | | | -0.037 |
| | | | (0.174) |
| Dependent variable mean | 4.224 | 4.224 | 4.224 |
| Year and village fixed effects | x | x | x |
| Observations | 1,080 | 1,080 | 1,080 |
| Villages | 69 | 69 | 69 |

***Note:*** *This table shows that (1) the presence of a VC of large clans is weakly associated with more levies villagers paid to the village government and that (2) the presence of village public investment projects is strongly correlated with a higher level of levies. The dependent variable is the log amount of average levies villagers paid to the village government in a particular year. The independent variables include a dummy variable indicating whether a VC came from the village's largest or second-largest clan, a dummy indicating any public investment projects during that year, and their interaction. Standard errors clustered at the village level are in parentheses. The sample is based on village-year observations of 69 villages, of which household level data are available, from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

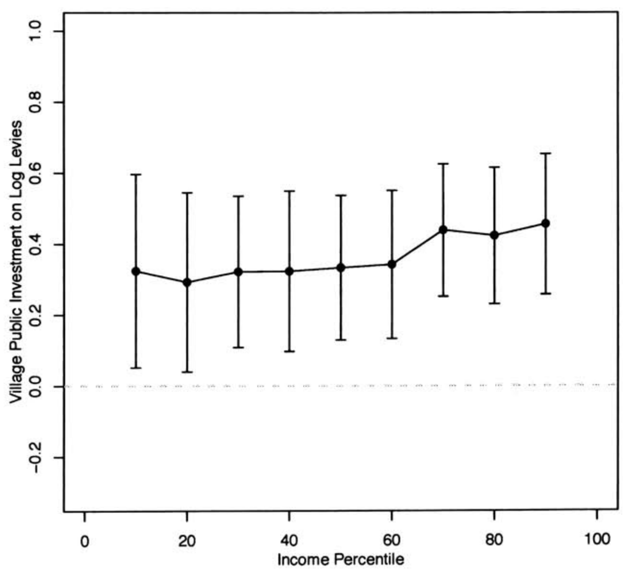
weakly associated with more levies. When a VC of large clans was in office, villagers on average paid 13.2 percent more levies a year. The estimated coefficient is not statistically significant, though, due to the large dispersion of the data. Column 2 regresses the log amount of levies on the public investment dummy when village and year fixed effects are controlled for. It shows that the amount of levies is highly correlated with the presence of public investment projects after time-invariant village heterogeneity and time-varying aggregate shocks are removed; the estimated coefficient is 0.304 and significant at the 1 percent level. In Column 3, we put in both dummies and their interaction.

The result is visualized in Figure 8a, which shows that no matter whether VCs of small

Figure 8. PUBLIC INVESTMENT AND LEVIES ON VILLAGERS




(a) Levies and public investment by VCs




(b) Levies and public investment by villagers' income

***Note:*** *Figure 8a shows that the amount of levies the average household in each village paid to the village committee during terms of VCs of small or large clans under two circumstances: when there was no public investment project during the year and when there was at least one project. Figure 8b shows the correlations between village public investment projects and the amount of levies households at specified income percentiles paid to the village committee. Village and year fixed effects are controlled for in both figures. The microlevel data in both figures come from 69 villages, a subset of the full sample.*

or large clans were in office, villagers paid more levies when there are public investment projects. On average, though, when VCs were from large clans, villagers paid higher levies to the village committee because higher frequencies of public investment projects were observed during the terms of VCs of large clans. Figure 8b shows the immediate distributive consequences of public investment on levies. Each dot is an estimated coefficient from a separate fixed effects regression using levies that households at a specified village income percentile paid to the village committee as the dependent variable and the public investment dummy as the independent variable. It shows that both the rich and poor in the villages paid extra levies when there were public projects. Note that the regressions presented in Figure 8 and Column 2 and 3 of Table 4 do not imply a causal relationship between the presence of public investment projects and the amount of levies villagers paid to the village committee, because both variables are likely results of the presence of VCs of large clans. They show, however, that to make a public investment project happen, a VC often needs to convince the majority of the villagers to pay for it. To the extent that VCs from large clans took up more investment projects than VCs from small clans, this allows us to conclude that large clans help VCs overcome the collective action problem.

Second, we investigate if there is any sign that informal institutions of large clans hold VCs accountable. We look at the amount of administrative expenditure of the village committee. If a VC is subject to close scrutiny when using public funds, non-productive administrative expenditure is most likely to be curbed. Previous studies have shown that electoral reforms in rural China caused a sharp decrease in the share of administrative expenditure in total expenditure of village committees (Wang and Yao 2007). Using the baseline specification (Equation 2) and both the share of administrative expenditure in total expenditure and the log amount of administrative expenditure as outcome variables, we show that VCs of large clans and administrative expenditure have almost zero correlation after village and year fixed effects are controlled for (Table 5).

Although we cannot rule out the possibility, we do not find strong evidence for the informal accountability mechanism.

Table 5. VCs of Large Clans and Administrative Expenditure

| | Share of administrative expenditure in total expenditure (1) FE | Log administrative expenditure (1,000 yuan) (2) FE |
|---|---|---|
| VC of large clans | 0.006 | 0.022 |
| | (0.014) | (0.073) |
| Dependent variable mean | 0.230 | 2.315 |
| Year and village fixed effects | x | x |
| Observations | 3,037 | 3,037 |
| Villages | 208 | 208 |

***Note:*** *This table shows that the association between VC of large clans and village administrative cost is close to zero after village and year fixed effects are controlled for. Standard errors clustered at the village level are in parentheses. In Column 1, the dependent variable is the share of administrative expenditure in total village expenditure in that year. In Column 2, the dependent variable is the log administrative expenditure (1,000 yuan). Both are from the NFS data. The independent variables are a dummy variable indicating whether a VC came from the village's largest or second-largest clan. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. Both regressions control for village and year fixed effects.*

## Alternative Explanations

In this section, we discuss two alternative explanations for the observed association between VCs of large clans and more public goods expenditure, including (1) superior ability of VCs of large clans, and (2) improvement of formal institutions.

First, do large clans select more competent leaders? Munshi and Rosenzweig (2010) find that in Indian parochial elections, castes with large population shares help select leaders with superior observed characteristics, such as providing more public goods. It

Table 6. Large Clans, VCs' Characteristics, and Village Public Investment

| | Log Public Investment (1,000 yuan) | | |
|---|---|---|---|
| | (1) FE | (2) FE | (3) FE |
| VC of large clans | 0.345 | 0.328 | 0.331 |
| | (0.127) | (0.129) | (0.129) |
| Years of education | -0.013 | -0.009 | -0.009 |
| | (0.021) | (0.024) | (0.025) |
| Age when running election | | -0.000 | -0.002 |
| | | (0.006) | (0.006) |
| CCP member | | 0.008 | 0.009 |
| | | (0.125) | (0.126) |
| Village cadre when running election | | -0.003 | 0.005 |
| | | (0.165) | (0.169) |
| Managerial jobs when running election | | 0.019 | -0.035 |
| | | (0.533) | (0.555) |
| Experience of running election | | 0.139 | 0.139 |
| | | (0.125) | (0.128) |
| Family background: poor peasant | | | -0.114 |
| | | | (0.153) |
| Denounced in the Culture Revolution (*pidou*) | | | 0.183 |
| | | | (0.325) |
| Dependent variable mean | 1.125 | 1.146 | 1.143 |
| Year and village fixed effects | x | x | x |
| Observations | 3,487 | 3,375 | 3,347 |
| Villages | 218 | 214 | 213 |

***Note:*** *This table shows that the association between a VC of large clans and village public investment is robust when we control for the VC's characteristics. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) in that year. The independent variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clan. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

is also possible that a successful entrepreneur from a large clan uses his or her resources and expertise to bring increased prosperity to the village.[30] To investigate these possibilities, we compile data of VCs' characteristics, including years of formal education, age,

[30] O'Brien (1994) reports that successful managers of collective enterprises were more likely to be trusted by villagers. Oi and Rozelle (2000) show that rural industrialization changed elites and other villagers' incentives to participate in grassroots politics.

administrative experience, experience of running businesses, CCP membership, historical family background, etc.[31] We control for these characteristics in the regressions. The results are shown in Table 6. The estimated coefficient of the VC of large clans remains almost unchanged. In fact, VCs' observed characteristics, such as education and administrative experience, do not seem to have any predictive power for the amount of public investment.[32] The evidence does not support that lineage groups in rural China helped select more competent leaders.

Another explanation is improvement in formal electoral institutions. As formal institutions improve, it is possible that elected leaders are more likely to implement policies catering to the median voter's interest, such as providing more public goods. It is also possible that under better formal institutions, officials elected into office have preferences that are more in line with preferences of the voters. These preferences might not have been captured by VCs' observed characteristics, but might be correlated with clans where VCs come from. Because our dataset has detailed information of electoral rules and procedures, including contested elections (an election is contested when there are more candidates than positions), open nomination, secret ballots, proxy voting, and moving ballot boxes, we can test if our main results are driven by changes of these indicators.[33] The results are shown in Table 7. As expected, our main finding is robust when we control for these indicators in the regressions. In fact, the institutional variations over time have very limited explanatory power for the variations in the amount of public investment. Moreover, the estimated coefficients of the VC dummy are slightly bigger in

[31] Historical family background was determined during the land reform in the 1950s by the local CCP authorities. After that, villagers from a poor peasant family background assumed most of the leadership position in the villages, as a legacy of the Communist revolution.

[32] In Appendix Table 15, we show that, compared with others, VCs of large clans did not have higher education or more administrative experience; after controlling for village and year fixed effects, we find that they appeared to be quite similar to the rest of the pool.

[33] Some of the indicators clearly suggest improvement in the electoral system, such as contested elections, open nomination and secret ballots. The impacts of proxy voting and moving ballot boxes are more ambiguous. They are supposed to increase the turnout of villagers, but they also create plenty of room for corruption and electoral frauds. Appendix Figure 12 shows the overtime changes of these indicators in our sample.

Table 7. VCs of Large Clans, Electoral Institutions, and Village Public Investment

| | Log Public Investment (1,000 yuan) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Full Sample | | | | | Contested Elections | Open Nomination |
| | (1) FE | (2) FE | (3) FE | (4) FE | (5) FE | (6) FE | (7) FE |
| VC of large clans | 0.369 | 0.364 | 0.370 | 0.368 | 0.368 | 0.432 | 0.377 |
| | (0.118) | (0.118) | (0.117) | (0.117) | (0.118) | (0.147) | (0.144) |
| Contested election | 0.001 | | | | | | |
| | (0.162) | | | | | | |
| Open nomination | | -0.146 | | | | | |
| | | (0.159) | | | | | |
| Secret ballot | | | 0.074 | | | | |
| | | | (0.157) | | | | |
| Proxy voting | | | | 0.039 | | | |
| | | | | (0.155) | | | |
| Moving ballot | | | | | 0.127 | | |
| | | | | | (0.129) | | |
| Dependent variable mean | 1.092 | 1.092 | 1.092 | 1.092 | 1.092 | 1.144 | 1.148 |
| Year and village fixed effects | x | x | x | x | x | x | x |
| Observations | 3,742 | 3,742 | 3,742 | 3,742 | 3,742 | 2,888 | 2,631 |
| Villages | 220 | 220 | 220 | 220 | 220 | 215 | 196 |

***Note:*** *This table shows that the association between a VC of large clans and village public investment is robust when we control for formal electoral institutions and procedures and when we use subsamples of contested elections and open nomination. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) in that year. The independent variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clan. Columns 1–5 use the full sample, which includes village-year observations from 1986 to 2005 after village elections were introduced. Columns 6 and 7 use sub-samples in which contested elections and open nomination were introduced, respectively. All regressions control for village and year fixed effects.*

two subsamples where contested elections and open nomination had been introduced, respectively.

## Conclusion

In the context of rural China, we find that informal institutions of lineage groups—rules and norms created and enforced by lineage groups—facilitate local public goods provision. Using fixed effects models as the main estimation strategy and a regression discontinuity design as a robustness check, we show that the presence of village chairpersons of large clans increased local public goods expenditure considerably. Such a relationship is stronger in villages where large clans persistently maintained lineage halls. Our finding is robust when we consider the roles of village party secretaries and village party organizations, as well as alternative explanations, such as superior observed characteristics of VCs of large clans and improved formal electoral institutions. This paper is among the first attempts to study the causal effect of informal institutions on governance outcomes.

We explore two possible channels: (1) informal institutions facilitate collective action of financing public goods among villagers, and (2) informal institutions hold VCs accountable to villagers. We show that the collective action channel is better supported by data. We find that villagers at almost all income percentiles paid extra levies to the village committee when there were public investment projects. However, we find little evidence that informal institutions held village officials accountable: on average the amount of administrative cost did not change when VCs of large clans were in office.

Two questions are not fully answered by this paper and require future research. The first is the possibility that large clans capture grassroots politics. The evidence presented in this paper suggests that large clans might have improved local governance in rural China in one specific aspect, namely, spending on public investment. However, it is possible that we do not measure outcomes that deteriorated because of clan power. For example, public goods expenditure as we have measured might have benefited members of large clans much more than the rest of the villagers, or VCs of large clans filled their pockets and those of their clan members' as they provided public goods. Large clans might collude with township officials to capture local politics as well. But because we do not have information on corruption or who used what public facilities, these consequences are not reflected by our study.

The fact that we do not observe clan capture might be due to that leaders of large clans were under tight control of the CCP. Although open nomination of candidates is the *de jure* procedure in village elections, the CCP, especially its organ at the township level, heavily intervenes in the nomination process. Moreover, as we discuss in the paper, the exercise of power of the VC is constantly checked by the CCP. Such a unique institutional arrangement may limit the generalizability of our finding. For example, in places where local leaders are not closely monitored and controlled by other parties or the upper-level government, informal institutions may enable leaders to extract rents from constituencies or target transfers to a narrow group of supporters.

The second question that requires more research is the co-evolution of formal and informal institutions. How do changes of formal institutions affect the functioning of informal institutions and how do political actors embedded in informal institutions respond to changes of incentives due to formal institutional changes? In this paper, we attempt to identify the effect of informal institutions in the context of rural democracy. Unfortunately, we cannot compare the effect *before and after* the introduction of elections due to data limitations. An equally interesting question is how the role of informal institutions has changed since the tax-and-fee reform deprived villages of their autonomous status of finance.

# Appendix

## Contents of the Appendix


1. Robustness checks for the main results

   Table 8: Baseline results using post-1995 and pre-2000 subsamples

   Table 9: Baseline results using a binary outcome variable

   Table 10: Baseline results with VC of the largest clan as the key independent variable

2. Informal institutions and clan size

   Table 11: VC of large clans and public investment: clan size

   Figure 9: The heterogeneous treatment effect w.r.t VC's clan size

   Figure 10: The effect of VCs of large clans on public investment: different thresholds

3. Clan cohesiveness and the role of the Communist Party

   Table 12: VCs of large clans, clan size and cohesiveness

   Table 13: Large clans and village party organizations

4. A regression discontinuity design: additional results

   Table 14 and Figure 11

5. Alternative explanations and additional results

   Table 15: Correlates between VC of large clans and other characteristics

   Figure 12: Evolution of electoral institutions

## Robustness Checks for the Main Results

In the main text, we only use observations in the post-election period. As a result, the panel is imbalanced. If the timing of the introduction of elections were correlated with the presence of a VC of large clans and public goods expenditure, the estimated coefficient of VC of large clans could be biased. O'Brien and Li (2006) report that regional governments did have concerns to introduce elections to villages that were dominated by one large lineage group. The governments were worried that the elected positions would be captured by the dominant clan, which would implement policies for the benefits of its members at the cost of others. To minimize potential biases caused by the onset of elections, we use a subsample of post-1995 observations and re-estimate the models. Since most villages already began elections in 1995, the panel is much more balanced.

Table 8 Columns 1–4 present the results. The estimates are slightly larger than the baseline results and remain statistically significant. Column 5–7 in the same table show that the estimates are stable when we drop observations after 2000, when the rural tax-and-fee reform started to be experimented within some regions. Note that we do not include village-specific time trends when using subsamples because the time series are too short, which results in highly singular variance-covariance matrix; however, the estimated coefficients of the VC dummies are always large and positive.

One might also be worried that our results are driven by a few extreme values. In Table 9, we replace the outcome variable with a binary indicator of whether there was any investment in a year and redo the exercises. The results show that on average a VC of large clans is associated with a 6–8 percent increase in the probability of public investment, or 25–35 percent of the dependent variable mean.

Table 10 shows that our main findings hold if we do not include the indicator of VC of the second-largest clan in regressions.

Table 8. VC of Large Clans and Village Public Investment: Subsamples

| | Log Public Investment (1,000 yuan) | | | | | |
|---|---|---|---|---|---|---|
| | After 1995 | | | Before 2000 | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | FE | FE | FE | FE | FE | FE |
| VC of the largest clan | 0.445 | 0.511 | 0.503 | 0.386 | 0.354 | 0.338 |
| | (0.215) | (0.205) | (0.210) | (0.173) | (0.178) | (0.188) |
| VC of the second-largest clar | 0.320 | 0.432 | 0.567 | 0.282 | 0.280 | 0.310 |
| | (0.243) | (0.256) | (0.272) | (0.159) | (0.161) | (0.169) |
| Dependent variable mean | 1.328 | 1.328 | 1.310 | 0.916 | 0.916 | 0.891 |
| Year fixed effects | x | x | x | x | x | x |
| Village fixed effects | x | x | x | x | x | x |
| Provincial linear trends | | x | x | | x | x |
| NFS controls | | | x | | | x |
| Observations | 2,317 | 2,317 | 2,220 | 2,644 | 2,644 | 2,448 |
| Villages | 220 | 220 | 217 | 217 | 217 | 206 |

***Note:*** *This table shows that the association between the presence of a VC of large clans and a larger amount of village public investment is robust in post-1995 and pre-2000 subsamples. Columns 1–3 use observations after 1995 while Columns 5–6 use observations before 2000. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) during that year. The independent variables are two dummy variables indicating whether a VC came from the village's largest or second-largest clan, respectively. The sample is based on village-year observations after village elections were introduced. All regressions control for both village and year fixed effects. In addition, Columns 2, 3, 5, and 6 control for provincial linear time trends. Columns 3 and 6 include five time-varying control variables from the NFS dataset, including average household size, arable land per capita, log income per capita, log village assets, and log village population.*

Table 9. VC of Large Clans and Village Public Investment: Binary Outcome

| | Binary Outcome: Any Pubic Investment | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | OLS | FE | FE | FE | FE | FE |
| VC from the largest clan | 0.059 | 0.082 | 0.078 | 0.077 | 0.074 | 0.094 |
| | (0.024) | (0.029) | (0.029) | (0.038) | (0.030) | (0.038) |
| VC of the second-largest clan | 0.040 | 0.060 | 0.062 | 0.062 | 0.060 | 0.060 |
| | (0.032) | (0.031) | (0.031) | (0.045) | (0.030) | (0.044) |
| Dependent variable mean | 0.231 | 0.231 | 0.231 | 0.231 | 0.228 | 0.257 |
| Year fixed effects | x | x | x | x | x | x |
| Village fixed effects | | x | x | x | x | x |
| Provincial linear trends | | | x | | x | x |
| Village linear trends | | | | x | | |
| NFS controls | | | | | x | x |
| Persons migrating out | | | | | | x |
| Taxes/fees to the upper-level government | | | | | | x |
| Transfers from the upper-level government | | | | | | x |
| Observations | 3,742 | 3,742 | 3,742 | 3,742 | 3,513 | 2,530 |
| Villages | 220 | 220 | 220 | 220 | 217 | 208 |

***Note:*** *This table shows that the presence of a VC of large clans is associated with a higher probability of a village public investment project. Standard errors clustered at the village level are in parentheses. The dependent variable is a dummy variable indicating whether there was any village investment during that year. The independent variables are two dummy variables indicating whether a VC came from the village's largest or second-largest clan, respectively. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. Column 1 controls for year fixed effects only; the rest control for both village and year fixed effects. In addition, Columns 3, 5, and 6 control for provincial linear time trends; Column 4 controls for village linear time trends; and Columns 5 and 6 include five time-varying control variables from the NFS dataset, including average household size, arable land per capita, log income per capita, log village assets, and log village population. Column 6 additionally controls for the number of persons migrating out of the village each year, log total taxes and fees the village committee handed over to the upper-level government and log transfers it received from the upper-level government, all of which are available after 1993 (the data for 1994 are interpolated).*

Table 10. VC of the Largest Clan and Village Public Investment

| | Log Public Investment (1,000 yuan) | | | | | |
|---|---|---|---|---|---|---|
| | (1) OLS | (2) FE | (3) FE | (4) FE | (5) FE | (6) FE |
| VC of the largest clan | 0.295 | 0.349 | 0.310 | 0.303 | 0.306 | 0.400 |
| | (0.121) | (0.145) | (0.144) | (0.180) | (0.152) | (0.193) |
| Dependent variable mean | 1.092 | 1.092 | 1.092 | 1.092 | 1.083 | 1.225 |
| Year fixed effects | x | x | x | x | x | x |
| Village fixed effects | | x | x | x | x | x |
| Provincial linear trends | | | x | | x | x |
| Village linear trends | | | | x | | |
| NFS controls | | | | | x | x |
| Persons migrating out | | | | | | x |
| Taxes to the upper-level government | | | | | | x |
| Transfers from the upper-level government | | | | | | x |
| Observations | 3,742 | 3,742 | 3,742 | 3,742 | 3,513 | 2,530 |
| Villages | 220 | 220 | 220 | 220 | 217 | 208 |

***Note:*** *This table shows that the presence of a VC of large clans is associated with a larger amount of village public investment. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) during that year. The independent variables is a dummy variable indicating whether a VC came from the village's largest clan. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. Column 1 controls for year fixed effects only; the rest control for both village and year fixed effects. In addition, Columns 3, 5, and 6 control for provincial linear time trends; Column 4 controls for village linear time trends; and Columns 5 and 6 include five time-varying control variables from the NFS dataset, including average household size, arable land per capita, log income per capita, log village assets, and log village population. Column 6 additionally controls for the number of persons migrating out of the village each year, log total taxes and fees the village committee handed over to the upper-level government and log transfers it received from the upper-level government, all of which are available after 1993 (the data for 1994 are interpolated).*

### Informal Institutions and Clan Size

In this section, we show that (1) our main results are robust when we control for the VC's clan size, (2) the effect of informal institutions, as we measure them, varies little across clans with different sizes, and (3) our results are robust when we use clan size (with different thresholds) as a measure of the strength of informal institutions. We also discuss why we think the rank order is a better measure for the clan's social power than the clan size.

**Does clan size matter?** First, we empirically test whether the magnitude of clan size matters. We directly incorporate both relative and absolute size of the VC's clan in two-way fixed-effect models. The results are reported in Table 11. In Column 1, the key independent variable is the relative size of the VC's clan, measured by the number of villagers in the VC's clan divided by the village's total population. The estimate is positive but not statistically significant. In Column 2, we additionally include the original rank order measure, in which case, we essentially treat the relative size of the VC's clan as a confounding factor. The estimated coefficient of the dummy variable is 0.438 and highly significant while the coefficient of relative clan size becomes negative and statistically insignificant. In Columns 3 and 4, we conduct similar tests but replace the relative size of the VC's clan by its absolute size (in 1,000 persons). The results are very similar. The estimated coefficient of the absolute size is positive but not significant. After we add the original rank order measure to the regression, the coefficient of the absolute size becomes almost zero, while the coefficient of the rank order measure is positive and highly significant. These results, taken at face value, show that once conditional on the rank order, the clan size has very limited explanatory power for the amount of public goods expenditure.

**Heterogeneous treatment effect.** Second, we want to know whether the effect of informal institutions on public goods expenditure is larger when the VC came from a

Table 11. VC of Large Clans and Public Investment: Clan Size

| | Log Public Investment (1,000 yuan) | | | |
|---|---|---|---|---|
| | (1) FE | (2) FE | (3) FE | (4) FE |
| Relative size of the VC's clan | 0.750 | -0.292 | | |
| | (0.426) | (0.564) | | |
| Absolute size of the VC's clan (1,000 persons) | | | 0.381 | -0.013 |
| | | | (0.279) | (0.335) |
| VC of the two largest clans | | 0.438 | | 0.355 |
| | | (0.160) | | (0.158) |
| Dependent variable mean | 1.092 | 1.092 | 1.077 | 1.077 |
| Year and village fixed effects | x | x | x | x |
| Observations | 3,742 | 3,742 | 3,530 | 3,530 |
| Villages | 220 | 220 | 208 | 208 |

***Note:*** *In this table, we explore the relationship between the VC's clan size, measured by the relative and absolute population share of the VC's clan, and the level of public investment. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) in that year. Note that we only record the size of the four largest clans (surnames) in a village; the size of other kinship groups is coded as 0. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

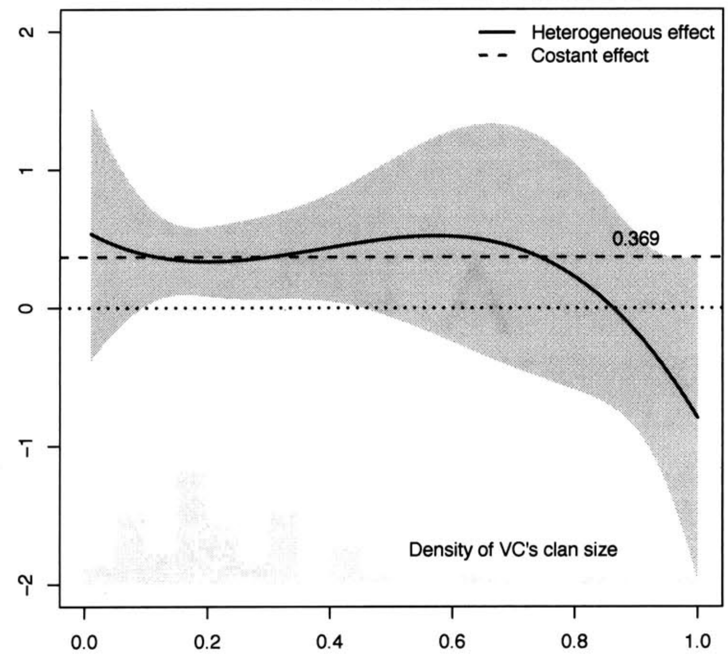
larger clan. In other words, we are interested in the heterogeneous treatment effect of VC of the two largest clans. We then interact the binary indicator VCs of large clans $D_{it}$ with a third-order polynomial of the size of the VC's clan:

$$y_{it} = \beta D_{it} + \gamma_1 D_{it} \times \omega_{it} + \gamma_2 D_{it} \times \omega_{it}^2 + \gamma_3 D_{it} \times \omega_{it}^3 + +\eta_i + \delta_t + \epsilon_{it}, \tag{3}$$

where $\omega_{it}$ is the population share of the VC's clan in village $i$ in year $t$ (we do not control for the level terms $\omega_{it}$, $\omega_{it}^2$, and $\omega_{it}^3$ because they are highly colinear with the interaction terms). The marginal effect of VCs of clans, therefore, is $(\beta+\gamma_1\omega_{it}+\gamma_2\omega_{it}^2+\gamma_3\omega_{it}^3)$. We are interested in whether the magnitude of the effect of informal institutions is dependent on the size of the VC's clan. The result is depicted in in Figure 9. Figure 9 shows that the effect of VC of large clans as measured by the rank order of VCs' clan size is relatively stable before the population share of the two largest clans reaches 75 percent.

In fact, they are close to the baseline estimate of 0.369 when a constant treatment effect is assumed. However, when the two largest clans consist of more than 75 percent of the village population, the estimates decline quickly and turn insignificant. This change occurs because (1) the number of villages with village-wide lineage groups is very small (as Figure 9 itself shows), and (2) there is simply not enough variation in the VC dummy since most of the VCs in these villages came from large clans.

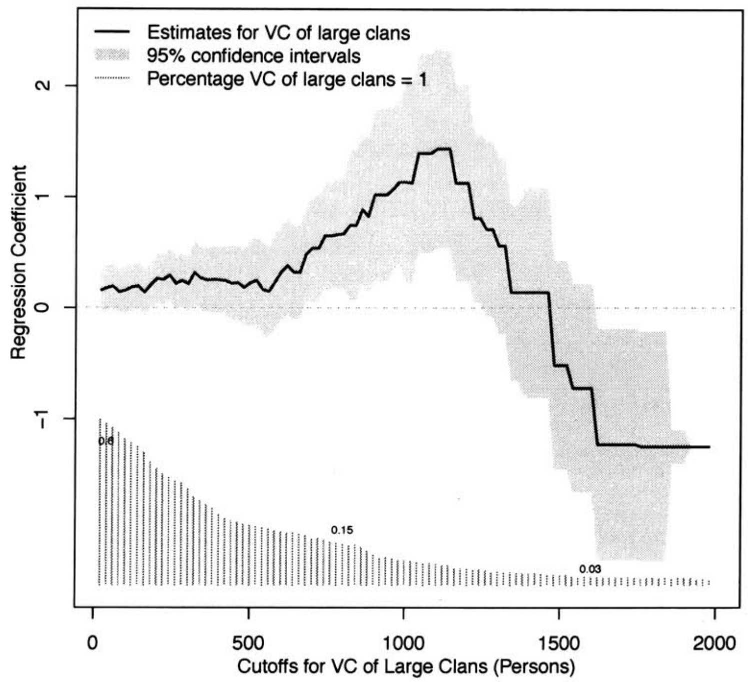
Figure 9. The Heterogenous Effect of VCs of Large Clans on Public Investment




***Note:*** *This figures shows the heterogeneous effect of VCs of large clans on the amount of public investment. The x-axis is the VC's clan size. The y-axis is the marginal effect of VC of large clans. The specification we use is shown in Equation 3.*

**Different thresholds.** In the main text, we mainly use the population rank order to measure a clan's social power (and hence, the strength of informal institutions associated with the VC's clan). In the following exercise, we measure the strength of lineage groups solely based on the number of people a clan has. In other words, if the size of a clan goes beyond a certain threshold, we code the group as a large clan, and estimate the effect of VC of large clans given the threshold. Because a threshold can be arbitrarily set, we try 100 thresholds with an interval of 20 persons between 0 to 2,000 persons (an average village in the period had around 1,500 villagers). The results of this analysis is shown in Figure 10. We find that the coefficient of VC of large clans is positive and statistically significant when the threshold is between 680 to 1240 persons, a large and reasonable interval. Moreover, if we exclude VCs from the third- and fourth-largest clans from VCs of large clans, the coefficient of VC of large clans is significant at almost all thresholds below 1240 persons. This means that even with the same group size, the largest and second-largest clans in a smaller village were fundamentally different from the third- and fourth-largest clans in a larger village in terms of social power.

Figure 10. THE EFFECT OF VCs FROM LARGE CLANS ON PUBLIC INVESTMENT: DIFFERENT THRESHOLDS




***Note:*** *This figure shows the estimated coefficients of VC of large clans using different threshold for large clans. For example, if the threshold is set at 500 persons, the dummy variable VC of large clans would equal one if the VC's clan consisted of more than 500 people and zero otherwise. The bars on the floor of the figure show the percentages of village-year observations when the variable VC of large clans equals one.*

Because of the large heterogeneities across the country, clans of the same absolute or relative size may have vastly different levels of social power. For example, a clan of 20 households in a socially fragmented village might be the largest clan of the village and thus more powerful than the largest clan in a village consisted of two clans with more or less equal sizes. Moreover, there can be much bigger measurement errors in the absolute or relative size of clans than in their population rank order, especially when we only took a snapshot in 2011. The size of a clan might have changed substantially over the 20-year period covered by our study, but the population rank order should be more stable. Measures of social cohesiveness, such as lineage halls and ceremonies can provide information about the intensity of within-clan social activities, but may not fully capture clans social power in the village. In the *Main Results* Section of the paper, indeed we see that it is the clan's social power that matters rather than its size.

In summary, we find that, the population rank order of clans is controlled for, the clan size has almost no predictive power for the amount of public goods expenditure. These results also indicate that the rank order of a VC's clan is a good proxy for the strength of informal institutions associated with the VC's clan.

### Clan Cohesiveness and the Role of Village Party Organizations

Figure 5 in the main text is based on the regression results reported in Table 12 Columns 1-3 with each column corresponding to a panel in the figure. In Column 4, when we put all three interaction terms in the regression, the coefficient of the interaction between the VC dummy and lineage halls remains large and significant. The coefficients of the other two interactions are negative but statistically insignificant.

Figure 6 in the main text is based on the regression results reported in Table 13 Columns 2-4 with each column corresponding to a panel in the figure. In Column 1, we only include the dummy variable indicating whether the VPS was from one of the two largest clans (VPS of large clans), as well as its interaction with VC of large clans. We find that the coefficient of VC of large clans is still large and statistically significant. The coefficient of VPS of the large clans is 0.249, slightly smaller than that of VCs of the largest clan, but statistically significant.

Table 12. VCs of Large Clans and Clan Cohesiveness

| | Log Public Investment (1,000 yuan) | | | |
|---|---|---|---|---|
| | (1) FE | (2) FE | (3) FE | (4) FE |
| VC of large clans | 0.433 | 0.277 | 0.180 | 0.301 |
| | (0.147) | (0.164) | (0.134) | (0.165) |
| × Combined size > 50% | -0.144 | | | -0.181 |
| | (0.242) | | | (0.256) |
| × Records of family trees | | 0.204 | | -0.107 |
| | | (0.261) | | (0.256) |
| × Lineage hall | | | 1.021 | 1.095 |
| | | | (0.331) | (0.338) |
| Dependent variable mean | 1.092 | 1.102 | 1.102 | 1.102 |
| Year and village fixed effects | x | x | x | x |
| Observations | 3,742 | 3,367 | 3,367 | 3,367 |
| Villages | 220 | 200 | 200 | 200 |

***Note:*** *This table shows that the association between a VC of large clans and village public investment is stronger in villages with more cohesive large clans, but it is not increasing in the VC's clan size. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) in that year. The independent variables are a dummy variable indicating whether a VC came from the village's largest or second-largest clan and its interactions with (1) whether the combined size of the two largest clans is above 50 percent, (2) whether any of the two largest clans had kept records of family trees, and (3) whether they had maintained any lineage halls since the beginning of the observed time periods. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

Table 13. LARGE CLAN LEADERS, VILLAGE PARTY ORGANIZATIONS, AND VILLAGE PUBLIC INVESTMENT

| | Log Public Investment (1,000 yuan) | | | |
|---|---|---|---|---|
| | (1) FE | (2) FE | (3) FE | (4) FE |
| VC of large clans | 0.509 | 0.581 | 0.580 | 0.473 |
| | (0.179) | (0.184) | (0.178) | (0.225) |
| VPS of large clans | 0.333 | 0.397 | 0.398 | 0.620 |
| | (0.183) | (0.189) | (0.191) | (0.258) |
| VC of large clans × VPS of large clans | -0.172 | -0.214 | -0.093 | -0.514 |
| | (0.274) | (0.288) | (0.335) | (0.417) |
| VC as the VPS ("one shoulder") | | 0.370 | | |
| | | (0.254) | | |
| × VC/VPS of large clans | | -0.476 | | |
| | | (0.367) | | |
| VC and VPS from the same clan | | | 0.215 | |
| | | | (0.214) | |
| × VC/VPS of large clans | | | -0.410 | |
| | | | (0.340) | |
| VC in the village party branch | | | | 0.219 |
| | | | | (0.220) |
| × VC of large clans | | | | 0.008 |
| | | | | (0.366) |
| × VPS of large clans | | | | -0.514 |
| | | | | (0.314) |
| × VC of large clans × VPS of large clans | | | | 0.506 |
| | | | | (0.505) |
| Dependent variable mean | 1.083 | 1.083 | 1.083 | 1.092 |
| Year and village fixed effects | x | x | x | x |
| Observations | 2,495 | 2,495 | 2,495 | 2,324 |
| Villages | 139 | 139 | 139 | 130 |

***Note:*** *This table shows that the association between a VC of large clans and village public investment is robust when we control for the roles of VPSs and village party organizations. The dependent variable is the log amount of village investment (1,000 yuan) in that year. The key independent variable is dummy variables indicating whether a VC came from the village's largest or second-largest clan, whether the VPS came from a village's largest or second-largest clan, and their interaction. In Addition, in Column 2, we control for whether the VC and VPS were the same person ("one-shoulder", or yijiantiao) and its interaction with VC of large clans. In Column 3, we control for whether the VC and VPS came from the same clan and their interactions with VC of large clans. In Column 4, we control for whether a VC was in the village party branch and its interactions with variables we included in Column 1. Standard errors clustered at the village level are in parentheses. The sample is based on village-year observations from 130-139 villages that report information on VPSs and village party organizations during the period of 1986-2005 after village elections were introduced. All regressions control for village and year fixed effects.*

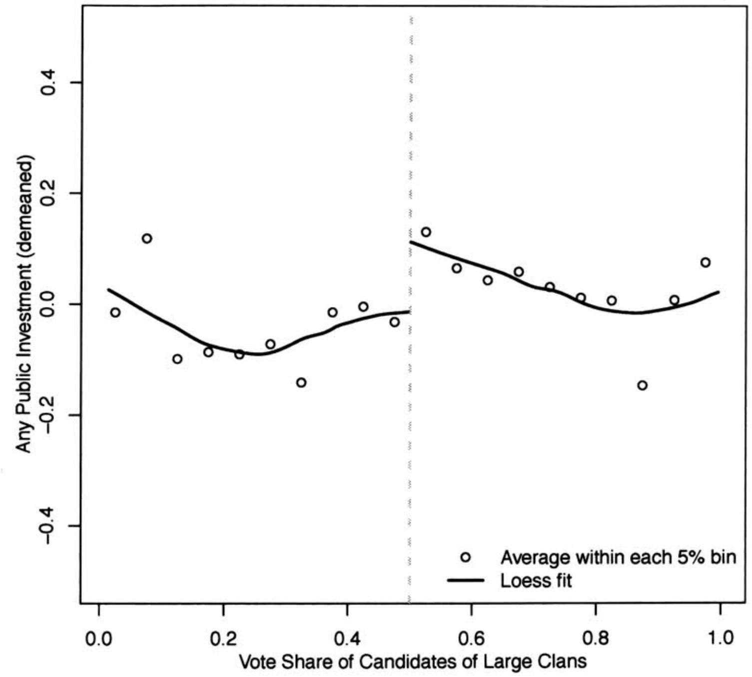
## Additional Results on the Regression Discontinuity Design:

Table 14. VC OF LARGE CLANS AND VILLAGE PUBLIC INVESTMENT: A REGRESSION DISCONTINUITY DESIGN

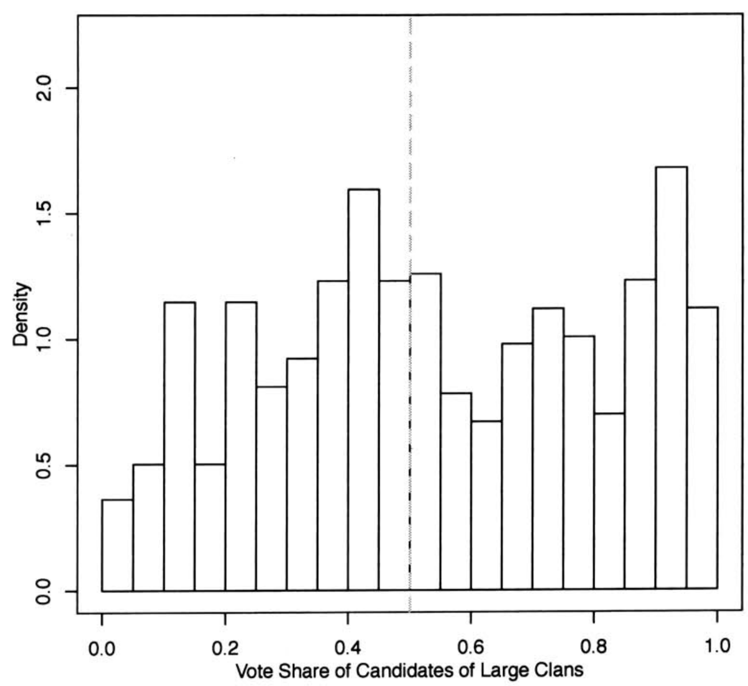
| *Panel A* | Log Investment (1,000 yuan) | | | | | |
|---|---|---|---|---|---|---|
| | All with #votes | Vote%≠{0 ,100} | Vote% [40,60] | Vote% [45,55] | 1st order poly. | 2nd order poly. |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | FE | FE | FE | FE | Loess | Loess |
| VC of large clans | 0.660 | 0.845 | 0.731 | 0.607 | 0.573 | 0.521 |
| | (0.189) | (0.355) | (0.847) | (0.820) | (0.301) | (0.435) |
| Dependent variable mean | 1.238 | 1.189 | 1.431 | 1.380 | 1.189 | 1.189 |
| Observations | 2,296 | 781 | 174 | 89 | 781 | 781 |
| Villages | 189 | 132 | 38 | 22 | 132 | 132 |
| *Panel B* | Binary Outcome: Any Investment | | | | | |
| | All with #votes | Vote%≠{0 ,100} | Vote% [40,60] | Vote% [45,55] | 1st order poly. | 2nd order poly. |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | FE | FE | FE | FE | Loess | Loess |
| VC of large clans | 0.125 | 0.172 | 0.170 | 0.166 | 0.123 | 0.124 |
| | (0.038) | (0.072) | (0.186) | (0.197) | (0.063) | (0.088) |
| Dependent variable mean | 0.257 | 0.251 | 0.310 | 0.315 | 0.251 | 0.251 |
| Observations | 2,296 | 781 | 174 | 89 | 781 | 781 |
| Villages | 189 | 132 | 38 | 22 | 132 | 132 |

***Note:*** *This table reports the estimates from an regression discontinuity design. In Panel A, the dependent variable is the log amount of village investment (1,000 yuan) in that year; in Panel B, it is a dummy variable indicating whether there was any village investment during that year. Both samples are based on village-year observations after village elections were introduced. The independent variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clan. Columns 1–4 report estimates from standard two-way fixed effects models. Standard errors clustered at the village level are in parentheses. In Column 1, observations without vote share data are dropped. In Column 2, observations in which a VC's vote share is either zero or one — neither the VC nor the runoff came from large clans (or both come from large clans) — are further dropped from the sample. Columns 5 and 6 limit the samples to relatively close elections, i.e. vote shares (%) of VCs of large clans are in the range of [40, 60] and [45, 55], respectively. Using the same sample as in Column 2, Columns 5 and 6 fit local linear regressions on both sides of the 50 percent cutoff and report the difference in the loess intercept estimates around the cutoff. Standard errors are produced by bootstraps of 1,000 times. The loess fits in Column 5 control for the level of the vote share (a first-order polynomial) while those in Column 6 control for the second-order polynomial. In Columns 5 and 6, observations are demeaned over time and within villages in advance to reduce dispersion and to account for aggregate shocks during the observed periods and time-invariant village heterogeneities.*

Figure 11. ROBUSTNESS CHECK: A REGRESSION DISCONTINUITY DESIGN (CONTINUED)




(a) Log public investment




(b) Histogram of the vote share

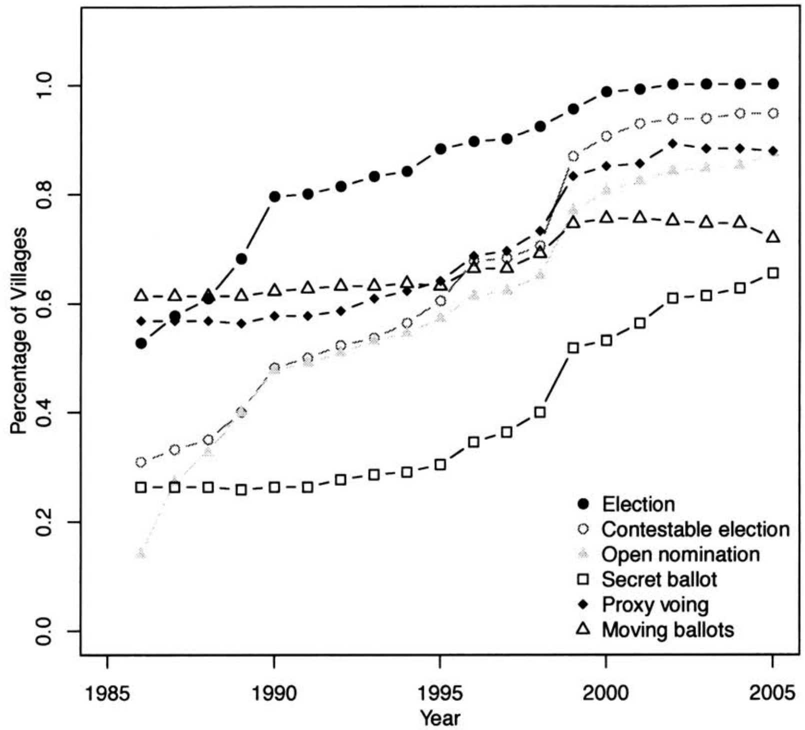
***Note:*** *Figure 11a shows the probability of any public investment projects within each 5 percent vote-share bin and two loess fits from locally linear regressions on both sides of the cutoff. Figure 11b plots the density of the vote-share of large-family candidates (values 0 and 1 not included).*

Table 15. LARGE CLANS AND VCS' CHARACTERISTICS

| *VC's characteristics* | Years of education (1) | Age when running election (2) | CCP member (3) | Village cadre when running election (4) | Managerial jobs when running election (5) | Experience of running election (6) | Family background: poor peasant (7) | Denounced in the Culture Revolution (*pidou*) (8) |
|---|---|---|---|---|---|---|---|---|
| VC of large clans | -0.145 | -0.163 | -0.052 | -0.033 | -0.001 | -0.040 | -0.028 | -0.020 |
| | (0.225) | (0.946) | (0.047) | (0.031) | (0.008) | (0.040) | (0.049) | (0.024) |
| Dependent variable mean | 6.39 | 41.6 | 0.75 | 0.56 | 0.02 | 0.71 | 0.79 | 0.05 |
| Year and village fixed effects | x | x | x | x | x | x | x | x |
| Observations | 1,210 | 1,203 | 1,195 | 1,209 | 1,209 | 1,205 | 1,213 | 1,203 |
| Villages | 218 | 219 | 216 | 218 | 218 | 216 | 219 | 216 |

***Note:*** *This table shows that VCs of large clans were not significantly different from those from small clans in terms of observed characteristics. Standard errors clustered at the village level are in parentheses. The dependent variables are observed characteristics of elected VCs. The independent variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clan. The sample is based on village-term observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

Figure 12. EVOLUTION OF ELECTORAL INSTITUTIONS IN THE SAMPLE VILLAGES




***Note:*** *This figure shows the changes of electoral rules and procedure from 1986 to 2005 in the sample villages.*

# Bibliography

Alesina, Alberto, R. Baqir and C.M. Hoxby. 2004. "Political Jurisdictions in Heterogeneous Communities." *Journal of Political Economy* 112(2):348–96.

Alesina, Alberto, Reza Baqir and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114(4):1243–84.

Banerjee, Abhijit and Lakshimi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95(4):1190–213.

Böröcz, Joszef. 2000. "Informality Rules." *East European Politics and Societies* 14(2):348–80.

Caughey, Devin and Jasjeet S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942-2008." *Political Analysis* 19(4):385–408.

Cohen, Myron L. 1990. "Lineage Organization in North China." *The Journal of Asian Studies* 49(3):509–34.

Collins, Kathleen. 2003. "The Political Role of Clans in Central Asia." *Comparative Politics* 35(2):171–90.

Collins, Kathleen. 2004. "The Logic of Clan Politics." *World Politics* 56(2):224–61.

Duara, Prasenjit. 1988. *Culture, Power, and the State: Rural North China, 1900-1942.* Stanford, CA: Stanford University Press.

Eggers, Andrew, Anthony Fowler, Jens Hainmueller, Andrew B. Hall and James M. Snyder. (forthcoming). "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races." *American Journal of Political Science* .

Fei, Hsiao-Tung. 1946. "Peasantry and Gentry: An Interpretation of Chinese Social Structure and Its Changes." *The American Journal of Sociology* 52(1):1–17.

Freedman, Maurice. 1958. *Lineage Organisation in Southeast China: Fukien and Kwangtung.* London: University of London and Athlone Press.

Freedman, Maurice. 1967. "Ancestor Worship: Two Aspects of the Chinese Case". In *Social Organization: Essays Presented to Raymond Firth*, ed. Maurice Freedman. Chicago: Aldine.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy M. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision." *American Political Science Review* 101(4):709–25.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy M. Weinstein. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action*. New York: Russell Sage Foundation.

Helmke, Gretchen and Steven Levitsky. 2004. "Informal Institutions and Comparative Politics: A Research Agenda." *Perspectives on Politics* 2(4):725–40.

Liu, Hui-chen Wang. 1959. *The Traditional Chinese Clan Rules*. New York: J J Augustin.

Luo, Renfu, Linxiu Zhang, Jikun Huang and Scott Rozelle. 2007. "Elections, Fiscal Reform and Public Goods Provision in Rural China." Working Paper, Freman Spogli Institute for International Studies, Stanford University, Stanford, CA.

Luo, Renfu, Linxiu Zhang, Jikun Huang and Scott Rozelle. 2010. "Village Elections, Public Goods Investment, and Pork Barrel Politics, Chinese-style." *Journal of Development Studies* 46(4):662–84.

Madsen, Richard. 1984. *Morality and Power in a Chinese Village*. Berkeley and Los Angeles, CA: University of California Press.

Manion, Melanie. 1996. "The Electoral Connection in the Chinese Countryside." *American Political Science Review* 90(4):736–48.

Manion, Melanie. 2006. "Democracy, Community, Trust: The Impact of Elections in Rural China." *Comparative Political Studies* 39(3):301–24.

Martinez-Bravo, Morica, Gerard Padró i Miquel, Nancy Qian and Yang Yao. 2011. "Do Local Elections in Non-Democracies Increase Accountability? Evidence from Rural China." NBER Working Paper 16948, National Bureau of Economic Research, Cambridge, MA.

Mattingly, Daniel. 2014. "One Party Rule of Divided Societies: Thwarting Ethnic Political Mobilization in China." Presented at the Annual Meeting of the Midwest Political Science Association, Chicago.

Munshi, Kaivan and Mark Rosenzweig. 2010. "Networks, Commitment, and Competence: Caste in Indian Local Politics." Mimeo, Brown University, Providence, RI.

Narayan, Deepa. 1999. *Bonds and Bridges: Social Capital and Poverty*. Washington, DC: World Bank.

O'Brien, Kevin. 1994. "Implementing Political Reform in Chinas Villages." *Comparative Politics Australian Journal of Chinese Affairs* 32:33–60.

O'Brien, Kevin and Lianjiang Li. 1999. "Selective Policy Implementation in Rural China." *Comparative Politics* 31(2):167–86.

O'Brien, Kevin and Lianjiang Li. 2000. "Accommodating 'Democracy' in a One-Party State: Introducing Village Elections in China." *The China Quarterly* 162:465–89.

O'Brien, Kevin and Lianjiang Li. 2006. *Rightful Resistance in Rural China.* Cambridge, UK: Cambridge University Press.

O'Brien, Kevin and Rongbin Han. 2009. "Path to Democracy? Assessing Village Elections in China." *Journal of Contemporary China* 18(60):359–78.

O'Donnell, Guillermo. 1996. "Another Institutionalization: Latin America and elsewhere." Kellog Institute for International Studies, Working Paper 222, Notre Dame, IN.

Oi, Jean. 1999. *Rural China Takes Off: Institutional Foundations of Economic Reform.* Berkeley: University of California Press.

Oi, Jean and Scott Rozelle. 2000. "Elections and Power: The Locus of Decision-Making in Chinese Villages." *The China Quarterly* 162(June):512–39.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, MA: Harvard University Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, UK: Cambridge University Press.

Pastor, Robert A. and Qingshan Tan. 2000. "The Meaning of China's Village Elections." *The China Quarterly* 162(June):490–512.

Perry, Elizabeth J. 2002. *Challenging the Mandate of Heaven: Social Protest and State Power in China.* New York: M. E. Sharpe.

Shen, Yan and Yang Yao. 2008. "Does Grassroots Democracy Reduce Income Inequality in China?" *Journal of Public Economics* 92(10-11):2182–98.

Shi, Tianjian. 1999. "Village Committee Elections in China: Institutionalist Tactics for Democracy." *World Politics* 51(3):385–412.

Sklar, Richard. 2004. "The Premise of Mixed Government in African Political Studies". In *Tradition and Politics: Indigenous Political Structures in Africa*, ed. Olufemi Vaughan. Trenton: Africa World Press.

Su, Fubing, Tao Ran, Xin Sun and Mingxing Liu. 2011. "Clans, Electoral Procedures and Voter Turnout: Evidence from Villagers Committee Elections in Transitional China." *Political Studies* 59:432–57.

Tan, Qingshan. 2006. *Village Elections: Democratizing the Chinese Countryside.* New York: Edwin Mellen Press.

Tan, Qingshan. 2009. "Building Democratic Infrastructure: Village Electoral Institutions." *Journal of Contemporary China* 18(60):411–20.

Tan, Qingshan. 2010. "Why Village Election Has Not Much Improved." *Journal of Chinese Political Science* (15):153–167.

Tsai, Lily L. 2007. "Solidary Groups, Informal Accountability, and Local Public Goods Provision in Rural China." *American Political Science Review* 101(2):355–72.

Unger, Jonathan. 2002. *The Transformation of Rural China.* New York: M.E. Sharpe.

Wang, Mingming. 1996. "Clans, the Society and the State: Rethinking Freedman's Theory." *Chinese Social Science Quarterly (Hong Kong)* 16.

Wang, Shuna and Yang Yao. 2007. "Grassroots Democracy and Local Governance: Evidence from Rural China." *World Development* 29(10):1635–49.

Watson, James. 1975. *Emigration and the Chinese Lineage.* Berkeley, CA: University of California Press.

Watson, James. 1982. "Chinese Kinship Reconsidered: Anthropological Perspectives on Historical Research." *The China Quarterly* 92:589–622.

Whiting, Susan. 1996. "Contract Incentives and Market Discipline in China's Rural Industrial Sector". In *Reforming Asian Socialism, the Growth of Market Institutions,* ed. John McMillan and Barry Naughton. Ann Arbor: University of Michigan Press.

Zhang, Jing. 2007. *Political Power on the Grassroots: Several Questions on Institutions in the Countryside (Jiceng Zhengquan: Xiangcun Zhidu Zhu Wenti).* Shanghai: Shanghai Century Publishing Group.

Zhang, Xiaobo, Shenggen Fan, Linxiu Zhang and Jikun Huang. 2004. "Local Governance and Public Goods Provision in Rural China." *Journal of Public Economics* 88:2857–71.