HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of **Government**

have examined a dissertation entitled

"Contributions to Casual Inference for Political Science"

presented by **Jens Hainmueller**

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature ______________________

Typed name: Prof. Gary King (Chair)

Signature ______________________

Typed name: Prof. Adam Glynn

Signature ______________________

Typed name: Prof. Michael Hiscox

Signature Alberto Abadie

Typed name: Prof. Alberto Abadie (HKS)

Date: **May 14, 2009**

Contributions to Causal Inference for Political Science

A dissertation presented
by

Jens Hainmueller

to
The Department of Government
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Government

Harvard University
Cambridge, Massachusetts
May 2009

Thesis advisor Author

Gary King Jens Hainmueller

Contributions to Causal Inference for Political Science

Abstract

This thesis presents five independent essays that advance causal inference in political science. It is divided into a methodological and an empirical part.

The methodological part presents a suite of statistical techniques, called synthetic inference methods, that allows researchers to construct control groups that more accurately resemble the treated units than is possible by commonly used methods of covariate adjustment. Synthetic inference is based on the idea of using weighted averages of control units to create so-called synthetic comparison units or comparison groups that resemble either a single treated unit or a group of treated units in all relevant characteristics that may confound a comparison. The thesis develops two variants of synthetic inference for empirical settings that are typically encountered in political science: comparative case studies (essay 1, with Alberto Abadie and Alexis Diamond) and cross-sectional studies (essay 2). Both methodological essays demonstrate the methods in real world applications and provide companion software for implementation.

The empirical part of the thesis presents three original empirical studies that contribute answers to previously unanswered causal questions about (a) the financial rewards to serving in Parliament (essay 3, with Andy Eggers), (b) the impact of foreign free media on the stability of authoritarian regimes (essay 4, with Holger Kern), and (c) the impact of economic concerns on public attitudes toward immigration (essay 5). The empirical essays advance the debates in these substantive fields by combining newly collected data and a design-based approach to causal inference. Design-based inference provides an effective strategy to identify valid control groups in settings where statistical control is insufficient since units potentially differ on more characteristics than can be measured and controlled for in a statistical model.

Contents

Title Page i
Abstract iii
Table of Contents v
List of Figures ix
List of Figures ix
List of Tables xi
List of Tables xi
Acknowledgments xiii
Dedication xv

1 Introduction **1**
1.1 Motivation and Structure 1
1.2 Summary of the Methodological Part 2
1.3 Summary of the Applied Part 4

2 Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program **7**
2.1 Introduction 7
2.2 Synthetic Control Methods for Comparative Case Studies 12
2.2.1 Comparative Case Studies 12
2.2.2 A Motivating Model 13
2.2.3 Implementation 16
2.2.4 Inference 18
2.3 Estimating the Effects of California's Proposition 99 21
2.3.1 Background 21
2.3.2 Data and Sample 25
2.3.3 Results 26
2.3.4 Inference about the effect of the California Tobacco Control Program 29
2.4 Conclusion 32
2.5 Appendix A: Data Sources 34

2.6 Appendix B: The Economic Impact of the German Reunification in West Germany . . . 35
2.7 Appendix C: Technical Details . . . 39
2.8 Tables for Chapter 2 . . . 42
2.9 Figures for Chapter 2 . . . 45

3 Synthetic Matching for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies 55
3.1 Introduction . . . 55
3.2 Setup of the Problem . . . 60
3.2.1 Causal Inference in Observational Studies with Binary Treatments 60
3.2.2 Propensity Score Weighting (PSW) . . . 63
3.2.3 Achieving Balance with Matching Methods . . . 64
3.3 Estimating Treatment Effects by Synthetic Matching . . . 67
3.3.1 Synthetic Matching . . . 67
3.3.2 Implementation . . . 71
3.3.3 Estimated Propensity Score as Base Weight . . . 73
3.3.4 Double Robust Estimation . . . 73
3.4 Monte Carlo Simulations . . . 74
3.4.1 Competing Estimators . . . 74
3.4.2 Monte Carlo Experiment I: EPBR Data . . . 76
3.4.3 Monte Carlo Experiment II: Non-EPBR Data . . . 78
3.4.4 Monte Carlo Experiment III: Non-EPBR Data with Various Designs . . . 79
3.5 Application: The LaLonde Dataset . . . 82
3.5.1 Data . . . 83
3.5.2 Balance Before Matching . . . 84
3.5.3 Balance After Matching . . . 84
3.5.4 Effect Estimates . . . 85
3.6 Conclusion . . . 85
3.7 Tables for Chapter 3 . . . 87
3.8 Figures for Chapter 3 . . . 96

4 The Value of Political Power: Estimating Returns to Office in Post-War British Politics 100
4.1 Introduction . . . 100
4.2 The Value of a Parliamentary Seat in Context . . . 105
4.3 The Wealth of Candidates to the House of Commons . . . 109
4.3.1 Data and Estimation Sample . . . 109
4.3.2 Wealth Distributions . . . 112
4.4 Estimating the effect of office on wealth . . . 113
4.4.1 Matching Estimates . . . 114

4.4.2 Regression Discontinuity Design Results 118
4.4.3 Robustness Tests for RD Estimation 122
4.5 Discussion 123
4.5.1 Did MPs Make their Money in Office or After Retiring? 124
4.5.2 How Did MPs Make Money In Office? 125
4.5.3 Why Did the Benefits of Office Differ by Party? 130
4.5.4 Did MPs' Outside Arrangements Affect Their Behavior? . . . 132
4.6 Conclusion 133
4.7 Tables for Chapter 4 135
4.8 Figures for Chapter 4 141

5 Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes 147
5.1 Introduction 147
5.2 The effects of foreign media exposure 149
5.3 Research design 153
5.3.1 Survey data 155
5.3.2 Causal Inference using instrumental variables 155
5.3.3 Covariates 161
5.3.4 Estimators 162
5.3.5 Outcome variables 163
5.4 Results 163
5.5 Exit visa applications 167
5.6 Historical evidence 168
5.7 Discussion 171
5.8 Robustness Supplement 174
5.8.1 LARF ordered probit 174
5.8.2 Subsample comparison 174
5.8.3 Replication I: Alternative survey 175
5.8.4 Replication II: Alternative control group 177
5.9 Tables for Chapter 5 178
5.10 Figures for Chapter 5 185

6 Economic Concerns and Attitudes Towards Immigration 189
6.1 Introduction 189
6.2 Economically Motivated Attitudes Toward Immigration 192
6.2.1 Labor Market Competition 192
6.2.2 The Fiscal Burden of Public Services 196
6.3 The Survey Experiment 200
6.3.1 Design 200
6.4 Empirical Test I: The Labor Market Competition Model 203
6.4.1 Skill Levels of Natives 203

6.4.2 Attitudes towards Highly and Low Skilled Immigrants and Natives' Skill Levels 204
6.4.3 Formal Tests of the Labor Market Competition Model 205
6.4.4 Results for Tests of the Labor Market Competition Model . . 207
6.5 Empirical Test II: The Fiscal Burden Model 210
6.5.1 Income and Fiscal Exposure to Immigration Across U.S. States 210
6.5.2 Attitudes towards Highly and Low Skilled Immigrants, Natives' Income, and Fiscal Exposure 213
6.5.3 Formal Tests of the Fiscal Burden Model 214
6.5.4 Results for the Tests of the Fiscal Burden Model 216
6.6 Conclusion 218
6.7 Appendix A: Theoretical Framework 220
6.8 Appendix B: Within-Groups Analysis 223
6.8.1 Stability of Attitudes and Skill Premium 224
6.8.2 Tests of the Labor Market Competition Model 225
6.8.3 Tests of The Fiscal Burden Model 226
6.9 Appendix C: Descriptive Statistics 228
6.10 Tables for Chapter 6 229
6.11 Figures for Chapter 6 232

Bibliography **240**

List of Figures

2.1 Trends in Per-Capita Cigarette Sales: California vs the Rest of the United States . . . 45
2.2 Trends in Per-Capita Cigarette Sales: California vs. synthetic California 46
2.3 Per-Capita Cigarette Sales Gap Between California and Synthetic California . . . 47
2.4 Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in all 38 Control States . . . 48
2.5 Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 34 Control States (Discards States with Pre-Proposition 99 MSPE Twenty Times Higher than California's) . . . 49
2.6 Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 29 Control States (Discards States with Pre-Proposition 99 MSPE Five Times Higher than California's) . . . 50
2.7 Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 19 Control States (Discards States with Pre-Proposition 99 MSPE Two Times Higher than California's) . . . 51
2.8 Ratio of Post-Proposition 99 MSPE and Pre-Proposition 99 MSPE: California and 38 Control States . . . 52
2.9 Trends in Per-Capita GDP: West Germany vs. Synthetic West Germany 53
2.10 Placebo in Time. Trends in Per-Capita GDP: West Germany vs. Synthetic West Germany . . . 54

3.1 Sample Designs for Second Monte Carlo Experiment: Density of True Propensity Score in Treatment and Control Group. . . . 96
3.2 Results for Second Monte Carlo Experiment ($r = 5$) . . . 97
3.3 The Effect of Synthetic Matching on Covariate Balance: QQ plots of Continuous Covariates . . . 98
3.4 The Effect of Synthetic Matching on Covariate Balance: Synthetic Matching versus Propensity Score Weighting. . . . 99

4.1 Fraction of Members of Parliament that declared Outside Interests 1975, 1990, and 2007 (fractions by party; dashed (solid) line decodes Labour (Conservatives)) . . . 141
4.2 Distributions of (Log) Wealth at Death by Party for Winning and Losing Candidates to the House of Commons 1950-1970 . . . 142
4.3 Covariate Balance Before and After Matching . . . 143
4.4 Regression Discontinuity Design: The Effect of Serving in the House of Commons on Wealth at Death . . . 144
4.5 Testing for Jumps at Non-discontinuity Points: Estimates for Conservative Candidates . . . 145
4.6 Outside Interests and Vote Attendance in the House of Commons 2005-2007 . . . 146

5.1 Over-the-air signal strength of West German television broadcasts . . . 185
5.2 District characteristics . . . 186
5.3 Over-the-air signal strength of West German television within the Dresden district . . . 187
5.4 Exit visa application rates and access to West German television for counties in the Dresden district . . . 188

6.1 Changes in Average Marginal State Income Tax Rate, Public Welfare Spending Per Capita, and Percent Foreign Born Population: 2004 to 1990 . . . 232
6.2 Support for Highly Skilled and Low Skilled Immigration . . . 233
6.3 Support for Highly Skilled and Low Skilled Immigration by Respondents' Skill Level . . . 233
6.4 Support for Highly Skilled and Low Skilled Immigration by Respondents' Skill Level . . . 234
6.5 Measures of Fiscal Exposure . . . 235
6.6 Attitudes Toward Highly Skilled and Low Skilled Immigration by Respondents' Income Level and Immigrant Fiscal Exposure of Respondents' State . . . 236
6.7 Support for Highly Skilled and Low Skilled Immigration by Respondents' Income Level and Immigrant Fiscal Exposure of Respondents' State . . . 237
6.8 Within-Groups Test: Support for Immigration by Respondents' Skill Level . . . 238
6.9 Within-Groups Test: Support for Immigration by Respondents' Income Level . . . 239

List of Tables

2.1 Cigarette Sales Predictor Means 42
2.2 State Weights in the Synthetic California 43
2.3 Economic Growth Predictor Means before the German Reunification 44
2.4 Country Weights in the Synthetic West Germany 44

3.1 Results for First Monte Carlo Experiment 87
3.2 Results for Second Monte Carlo Experiment 88
3.3 Results for Second Monte Carlo Experiment (N=300) 89
3.4 Results for Second Monte Carlo Experiment (N=600) 90
3.5 Results for Second Monte Carlo Experiment (N=1,500) 91
3.6 Covariance Balance Before Matching: Treatment versus Control Group 92
3.7 Covariance Balance after Synthetic Matching: Treatment versus Matched Control Group 93
3.8 Estimates for Average Treatment Effect for the Treated 94
3.9 Covariance Balance after Propensity Score Weighting: Treatment versus Matched Control Group 95

4.1 Gross Wealth at Death (Real 2007 GBP) for Competitive Candidates Who Ran for the House of Commons Between 1950-1970 (Estimation Sample) 135
4.2 Characteristics of Competitive Candidates Who Ran for the House of Commons Between 1950-1970 (Estimation Sample) 136
4.3 Matching Estimates: The Effect of Serving in the House of Commons on (Log) Wealth at Death 136
4.4 Regression Discontinuity Design Results: The Effect of Serving in the House of Commons on (Log) Wealth at Death 137
4.5 The Effect of Serving on Placebo Outcomes 138
4.6 Characteristics of the Political Careers of Members of Parliament (Estimation Sample) 139
4.7 The Correlates of Wealth: Estimates for Conservative MPs (Estimation Sample) 140

5.1 West German television exposure and place of residence 178
5.2 Outcome variables 178
5.3 Effect of West German television exposure on regime support 179
5.4 Deterrence and escapism 180
5.5 Exit visa application rates in 1988 180
5.6 Effect of West German television exposure on regime support: First differences from ordered probit LARF estimator 181
5.7 Effect of West German television exposure on regime support (subsamples) 182
5.8 Effect of West German television exposure on regime support: Replication with 1984 survey 183
5.9 Effect of West German television exposure on regime support: Replication with 1984 survey and Greifswald respondents as control group 184

6.1 Individual Support for Highly Skilled and Low Skilled Immigration - Test of the Labour Market Competition Model 229
6.2 Individual Support Highly Skilled and Low Skilled Immigration - Test of the Fiscal Burden Model 230
6.3 Split-Sample Cross-Over Design for Within-Groups Test 231
6.4 Mean Support for Immigration by Module and Immigrants' Skill Type 231

Acknowledgments

I would like to genuinely thank everybody who has helped and supported me during my graduate studies at Harvard. I have been very privileged to study at this great institution and to learn from professors that have made key contributions to the causal inference revolution in recent decades.

I cannot emphasize enough how thankful I am to have Gary King as my advisor. He is undoubtedly the most supportive advisor anyone could ask for (reread this sentence I mean it). I am still waiting for the day when it will take him longer than a few minutes to respond to my emails. The time and energy that he has devoted to support me is truly remarkable. I hope that I will be able to provide the same support for my own students in the future - he has been a shining example. I am also grateful that he offered me the opportunity to work as his teaching fellow for Gov 2001. This has been an invaluable and very instructive experience. Finally, I like to congratulate Gary on creating the collaborative and high-powered research environment at the Institute for Quantitative Social Science which has been an outstanding resource for methodological innovation and financial support.

I am deeply indebted to Alberto Abadie for sparking my interest in causal inference when I first took his program evaluation class. Although it took me a while to square the causal inference perspective with the techniques that I had learned in previous classes, it fundamentally transformed my thinking about statistical inference and study design. Throughout my years at Harvard, Alberto's rigor, persistence, and his constant flow of brilliant ideas has taught me a lot. I also like to thank him for taking me on as his teaching fellow and as a coauthor on a paper that is included in this dissertation. Learning from and working with Alberto was among the very best experiences of my graduate studies.

Michael Hiscox has been a close mentor and wonderful collaborator. I still remember well how we first met at the old Weatherhead Center. Since then, Michael has crucially influenced my work and thinking. He has inspired many new ideas and taught me a great deal about political economy research. I am very grateful for the wise guidance and excellent support that he has provided over the years – this means a lot to me and if I never met Michael I would not be where I am today. So thank you!

I thank Adam Glynn for letting me learn from him as his teaching assistant and for many helpful suggestions, insights, and comments that he has provided to my studies. I thank Jim Snyder for many helpful discussions and for organizing MIT's Political Economy Breakfast, which has been an exceptionally open and friendly forum for many fruitful debates and exchanges; Don Rubin and Jas Shekon for teaching me a great deal about causal inference, Beth Simmons for being supportive when I first joined the Government Department, and Thom Wall for all the administrative help that he has provided.

Very special thanks goes to my friend Andy Eggers, who has been a true companion and brilliant collaborator throughout the madness that is entailed in the pursuit of a PhD. His inspiring friendship has made the good times more shining and lessened the bad times by dividing and sharing it. I am also thankful to my friend and collaborator Alexis Diamond with whom I've shared many fascinating exchanges. Jason Lakin deserves special mention for bringing the "wise men" together for frequent discussions and I am thankful to my friends Sebastian Bauhoff and Dilyan Donchev for the good times we shared and the valuable feedback they have provided to my work. I also express my gratitude to my loving parents, Hiltrud and Bernd, who have nurtured my interest in politics and have always believed in me, as well as my siblings Anke and Moritz for their moral support. I cherish your love.

Finally, I am most deeply indebted to the love of my life and fiancée Jenny Suckale, who has been my shining light from the moment I started my PhD studies. During all these years, she has given me life's greatest happiness and love and I dearly hope that she will stay with me forever.

Dedicated to my fiancée Jenny,
my parents Hiltrud and Bernd,
my siblings Anke and Moritz,
and the Raubtierzoo.

Chapter 1

Introduction

1.1 Motivation and Structure

The dissertation consists of five self-contained essays within the rubric of causal inference for political science. The goal in causal inference is to estimate the effect of treatments, interventions, or policies within the framework of potential outcomes that contrasts similar units under active treatment or control (see for example Holland (1986)). Many questions in political science are causal, but modern methods of causal inference are still rarely used in political science research despite the fact that such methods have gained considerable grounds in other disciplines such as economics, epidemiology, and program evaluation. Sometimes there is even a latent scepticism that these methods cannot be fruitfully applied to study the political realm where causal inference is often exceptionally difficult.

For causal inference to succeed the researcher needs to find a treatment and a control group that are equivalent or "balanced" in observed and unobserved background characteristics so that differences in outcomes can be attributed to the effect of the treatment and isolated from pre-existing differences. Since Fisher (1925) we know that this is straightforward in experiments where the researcher can construct balanced groups by random assignment of the treatment. But in political science we often have to resort to observational studies because we cannot randomly assign our treatments or it would be even unethical to do so. In observational studies extremely careful study design is required to identify causal effects. The researcher needs to ask the vexing question: "To whom should the treated be compared?" The key challenge is to find a valid control group that accurately resembles the treatment group, whether the treatment under consideration be political office, the passage of a new law, or regime type. There can hardly be any credible causal inference without a valid control group. However, the methods currently used in political science to construct or find valid control groups are still fairly limited and the field of political methodology has paid insufficient attention to assist applied researchers in this crucial task.

This thesis contributes to the causal inference literature in political science both methodologically and empirically. It is therefore divided into two parts. The first part is primarily methodological and lays out the statistical innovations (chapters 1-2). It develops methods to statistically construct better control groups in two typical political science settings, namely comparative case studies and cross-sectional studies. The second part of the thesis is primarily empirical, and presents three original studies that use newly collected data and design-based inference to answer three previously unanswered causal questions that arise in political science debates. Beyond contributing new empirical knowledge to these debates, the empirical essays demonstrate how "controlling by design" provides an effective empirical strategy to identify valid control groups in political science settings where model-based control is insufficient because units potentially differ on more characteristics than can be measured and accounted for in a statistical model.

1.2 Summary of the Methodological Part

The methodological part of the thesis presents a suite of statistical techniques, called synthetic inference methods, that allows researchers to construct control groups that more accurately resemble the treated units than is possible by commonly used methods of covariate adjustment. Synthetic inference is based on the idea of using weighted averages of control units to create so-called synthetic comparison units or comparison groups that resemble either a single treated unit or a group of treated units in all relevant characteristics that may confound a comparison. We present two variants of synthetic inference for typical political science settings, namely comparative case studies and cross-sectional studies.

The first essay presents the synthetic control method for comparative case studies building on the pioneering paper by Abadie & Gardeazabal (2003). We derive an estimator that identifies the effect of aggregate level events or interventions (such as the passage of a law, an economic shock, etc.) on a single treated entity (such as a region, a country, or a school, etc.) in a way that is consistent with the traditional comparative case study design but more general than the conditions under which traditional linear panel data or difference-in-differences estimators are valid. It offers an appealing data driven rationale to tackle the classic problem of case selection. We also develop new methods for permutation inference in the comparative case study framework.

In the second essay we take synthetic inference to the cross-sectional setting and develop synthetic matching as a technique to create balanced samples for the estimation of treatment effects under the assumption of selection on observables. Synthetic matching allows researchers to calibrate a set of unit specific weights so that the reweighted control group exactly matches the treatment group on a potentially large set of pre-specified balance constraints. We show that these balance improvements translate into lower bias and higher efficiency for the estimation of treatment effects in finite samples compared to other adjustment methods commonly used in this context.

Both methodological papers emphasize the applicability of the proposed tools for research in political science, demonstrate the strength of the methods in real world empirical applications, and offer companion software.[1]

[1] One of the two software packages is published as Abadie, Diamond & Hainmueller (2009).

1.3 Summary of the Applied Part

The applied part of the thesis presents three original empirical studies that answer three previously unanswered causal questions arising from political science debates. Each study demonstrates how carefully crafted research designs combined with newly collected data can identify causal quantities of interest that directly speak to important theoretical debates in the relevant fields.

The first empirical study presents the first estimates of the returns to political office, an often cited quantity of interest in the political economy literature, in Post-War British politics. We like to know how much Members of Parliament (MPs) gained financially as a result of serving in the legislature. Since political office is not randomly assigned, the key challenge is to find a valid control group of people that are similar to MPs, but did not serve in Parliament. We solve this identification problem by careful choice of the research design. We track down the wealth of narrowly defeated candidates that we can use in a regression discontinuity design that exploits the quasi random assignment to office in close races (as in Lee (2008) or Hainmueller & Kern (2008)). If a race is decided by a razor thin margin the outcome is essentially random and thus the groups of marginally winning and marginally loosing candidates are essentially identical in terms of their characteristics allowing us to attribute any differences in wealth to the effects of serving in parliament. We find that serving in Parliament roughly doubled the wealth at death of Conservative MPs but had no discernible effect on the wealth of Labour MPs.

The second empirical study re-examines the common claim in democratization theory that penetration by foreign free media undermines authoritarian rule. No reliable micro-level evidence on this topic exists given that independent survey research is rarely possible in authoritarian regimes. Moreover, causal inference is rather complicated because people self-select into media consumption and thus a comparison of those who chose to consume foreign media with those that do not is not informative about the effect of foreign media on regime support (just imagine staunch regime supporters refuse to watch foreign television). The key challenge is to find two identical groups of people only one of which is exposed to foreign media. We solve this thorny identification problem using a case study of the impact of West German television on political attitudes in communist East Germany, a "most likely" case for democratization theory given that the potential effect of foreign free media was not diluted by language or cultural barriers. We find a valid control group by exploiting variation in media exposure that arose from East Germany's topography. While most East Germans enjoyed access to West German television some where cut off from the signal because of mountainous terrain. Our results show that East Germans who could tune in to West German television became more, and not less, satisfied with the East German regime, compared to those who could not. In contrast to what theories of democratization predict, the narcotizing effect of television served to stabilize rather than to undermine communist rule.

The third empirical study presents the first experimental test of the two major theories that predict how economic concerns generate anti-immigrant sentiment among native citizens: concerns about labor market competition and fears about the fiscal burden on public services. Both theories begin with a general equilibrium model and derive predictions about how native citizens who own different types of productive factors, and who have different levels of income, will differ in their views regarding highly skilled and low skilled immigration. We provide direct tests of both models of attitude formation using an original survey experiment embedded in a nationwide U.S. survey. The labor market competition model predicts that natives will be most opposed to immigrants who have similar skill levels to their own. We find instead that both low skilled and highly skilled natives strongly prefer highly skilled immigrants over low skilled immigrants, and this preference is not decreasing in natives' skill levels. The fiscal burden model anticipates that rich natives oppose low skilled immigration more than poor natives, and that this gap is larger in states with greater fiscal exposure (in terms of immigrant access to public services). We find instead that rich and poor natives are equally opposed to low skilled immigration, and rich natives are actually less opposed to low skilled immigration in states with more fiscal exposure than they are elsewhere. We do find that poor natives are more opposed to low skilled immigration in states with greater fiscal exposure than elsewhere, suggesting that concerns about access to or overcrowding of public services contributes to anti-immigrant attitudes.

Chapter 2

Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program

2.1 Introduction

Economists and other social scientists are often interested in the effects of events or policy interventions that take place at an aggregate level and affect aggregate entities, such as firms, schools, or geographic or administrative areas (countries, regions, cities, etc.).[1] To estimate the effects of these events or interventions, researchers often use comparative case studies. In comparative case studies, researchers estimate the evolution of aggregate outcomes (such as mortality rates, average income, crime rates, etc.) for a unit affected by a particular occurrence of the event or intervention of interest and compare it to the evolution of the same aggregates estimated for some control group of unaffected units. Card (1990) studies the impact of the 1980 Mariel Boatlift, a large and sudden Cuban migratory influx in Miami, using other cities in the southern United States as a comparison group. In a well-known study of the effects of minimum wages on employment, Card and Krueger (1994) compare the evolution of employment in fast-food restaurants in New Jersey and its neighboring state Pennsylvania around the time of an increase in New Jersey's minimum wage. Abadie and Gardeazabal (2003) estimate the effects of the terrorist conflict in the Basque Country on the Basque economy using other Spanish regions as a comparison group.

Comparing the evolution of an aggregate outcome (e.g., state-level crime rate) between a unit affected by the event or intervention of interest and a set of unaffected units requires only aggregate data, which are often available. However, when data are not available at the same level of aggregation as the outcome of interest, information on a sample of disaggregated units can sometimes be used to estimate the aggregate outcomes of interest (like in Card, 1990, and Card and Krueger, 1994).[2]

---

[1]This paper is co-authored with Alberto Abadie, John F. Kennedy School of Government, and Alexis Diamond, Political Economy and Government. We thank Joshua Angrist, Jake Bowers, Javier Gardeazabal, Bryan Graham, Dan Hopkins, Rustam Ibragimov, Guido Imbens, Paul Rosenbaum, Don Rubin, Jas Sekhon, Gary Solon, Chris Winship, and seminar participants at Harvard, Syracuse, Uppsala, the NBER Labor Meetings, and the 2006 APSA Meetings in Philadelphia for helpful comments. Funding for this research was generously provided by NSF grant SES-0350645 (Abadie). This paper was awarded the 2007 Gosnell Prize for the best work in methods presented at any political science conference during the preceding year. Companion software developed by the authors (Synth package for MATLAB, R, and Stata) is available at the authors' website and published as Abadie et al. (2009).

[2]Card (1990) uses individual-level data from the U.S. Current Population Survey to estimate the unemployment rates of native workers in Miami and a group of comparison cities before and after the arrival of the Mariel expatriates to Miami in 1980. Card and Krueger (1994) use a telephone survey of fast-food restaurants in New Jersey and Pennsylvania to estimate average wages and employment in the fast-food industry in those two states around the time of the increase in minimum wage in

Given the widespread availability of aggregate/macro data (for example, at the school, city, or region level), and the fact that many policy interventions and events of interest in the social sciences take place at an aggregate level, comparative case study research has broad potential. However, comparative case study research remains limited in economics and other social sciences, perhaps because its empirical implementation is subject to two elusive problems. First, in comparative case studies there is typically some degree of ambiguity about how comparison units are chosen. Researchers often select comparison groups on the basis of subjective measures of affinity between affected and unaffected units. Second, comparative case studies typically employ data on a sample of disaggregated units and inferential techniques that measure *only* uncertainty about the aggregate values of the data in the population. Uncertainty about the values of aggregate variables can be eliminated completely if aggregate data are available. However, the availability of aggregate data does not imply that the effect of the event or intervention of interest can be estimated without error. Even if aggregate data are employed, there remains uncertainty about the ability of the control group to reproduce the counterfactual outcome trajectory that the affected units would have experienced in the absence of the intervention or event of interest. This type of uncertainty is not reflected by the standard errors constructed with traditional inferential techniques for comparative case studies.

This article addresses current methodological shortcomings of case study analysis. We advocate the use of data-driven procedures to construct suitable comparison groups, as in Abadie and Gardeazabal (2003). Data-driven procedures reduce discretion in the choice of the comparison control units, forcing researchers to demonstrate the affinities between the affected and unaffected units using observed quantifiable characteristics. In practice, however, it is often difficult to find a single unexposed unit that approximates the most relevant characteristics of the unit(s) exposed to the event of interest. The idea behind the synthetic control approach is that a combination of regions often provides a better comparison for the region exposed to the intervention than any single region alone. For example, in their study of the eco-

New Jersey in 1992.

nomic impact of terrorism in the Basque Country, Abadie and Gardeazabal (2003) use a combination of two Spanish regions to approximate the economic growth that the Basque Country would have experienced in the absence of terrorism. Card (1990) implicitly uses a combination of cities in the southern United States to approximate the evolution that the Miami labor market would have experienced in the absence of the Mariel Boatlift.

Relative to traditional regression methods, transparency and safeguard against extrapolation are two attractive features of the synthetic control method. Because a synthetic control is a weighted average of the available control units, the synthetic control method makes explicit (1) the relative contribution of each control unit to the counterfactual of interest; and (2) the similarities (or lack thereof) between the unit affected by the event or intervention of interest and the synthetic control, in terms of pre-intervention outcomes and other predictors of post-intervention outcomes. Because the weights can be restricted to be positive and sum to one, the synthetic control method provides a safeguard against extrapolation.

In addition, because the choice of a synthetic control does not require access to post-intervention outcomes, the synthetic control method allows researchers to decide on study design without knowing how those decisions will affect the conclusions of their studies. Rubin (2001) and others have advocated that the ability to make decisions on research design while remaining blind to how each particular decision affects the conclusions of the study is an important device for promoting research honesty in observational studies.

We describe a simple econometric model that justifies the synthetic control approach. The model extends the traditional linear panel data (difference-in-differences) framework, allowing that the effects on unobserved variables on the outcome vary with time. In addition, we propose new methods that allow researchers to perform inferential exercises about the effects of the event or intervention of interest that are valid regardless of the number of available comparison units, the number of available time periods, and whether aggregate or individual data are used for the analysis.

We apply the synthetic control method to study the effects of California's Proposition 99, a large-scale tobacco control program implemented in California in 1988. We demonstrate that following the passage of Proposition 99 tobacco consumption fell markedly in California relative to a comparable synthetic control region. We estimate that, by the year 2000, annual per-capita cigarette sales in California were about 26 packs lower than what they would have been in the absence of Proposition 99. Using new inferential methods proposed in this paper, we demonstrate the statistical significance of our estimates.

Cross-country regressions are often criticized because they put countries side-by-side regardless of whether they have similar or radically different characteristics (see, for example, Temple, 1999). The synthetic control method provides an appealing data-driven procedure to select comparison groups for the study of the effects of events or interventions that take place at the level of a country. To illustrate the application of the techniques proposed in this article to cross-country data, we include an appendix where we use the synthetic control method to estimate the impact of the 1990 German re-unification on the West German economy.

The rest of the article is organized as follows. Section 2.2 describes the main ideas behind the synthetic control approach to comparative case studies of aggregate events. In section 2.3 we apply synthetic control methods to estimate the effect of California's Proposition 99. Section 2.4 concludes. Appendix A lists the data sources for the application in section 2.3. Appendix B contains the application of the synthetic control method to the study of the economic effects of the German reunification. Appendix C contains technical details.

2.2 Synthetic Control Methods for Comparative Case Studies

2.2.1 Comparative Case Studies

Case studies focus on particular occurrences of the events or interventions of interest. Often, the motivation behind case studies is to detect the effects of an event or policy intervention on some outcome of interest by focusing on a particular instance in which the magnitude of the event or intervention is large relative to other determinants of the outcome, or in which identification of the effects of interest is facilitated by some other characteristic of the intervention. For example, in his classic study of the economic impact of immigration, Card (1990) analyzes the behavior of the Miami labor market in the wake of the 1980 Mariel Boatlift, when the Mariel immigrants increased the size of the labor force in Miami by 7 percent in a matter of a few months. In comparative case studies, researchers compare units affected by the event or intervention of interest to a group of unaffected units. Therefore, comparative case studies are only feasible when some units are exposed and others are not (or when their levels of exposure differ notably).[3]

To simplify the exposition, we proceed as if only one unit or region is subject to the intervention of interest.[4] In addition, we adopt the terms "region" or "unit" and "intervention" or "treatment", which can be substituted for "country", "state", "city", etc. and "event", "shock", "law", etc., respectively for specific applications.

[3]In comparative case studies, the main emphasis is sometimes on identification of the impact of the particular event or intervention on hand (internal validity), at the cost of limited immediate generalizability to other settings (external validity). In other instances, as in the Card and Krueger (1994) study on the employment effects of a minimum wage raise, cases studies are used to test hypotheses previously derived from theoretical models.

[4]Otherwise, we could first aggregate the data from the regions exposed to the intervention.

2.2.2 A Motivating Model

The following simple model provides a rationale for the use of synthetic control methods in comparative case study research. Suppose that we observe $J+1$ regions. Without loss of generality, suppose also that only the first region is exposed to the intervention of interest, so that we have J remaining regions as potential controls. Also without loss of generality and to simplify notation, we assume that the first region is uninterruptedly exposed to the intervention of interest after some initial intervention period.

Let Y_{it}^N be the outcome that would be observed for region i at time t in the absence of the intervention, for units $i = 1, \ldots, J+1$, and time periods $t = 1, \ldots, T$. Let T_0 be number of pre-intervention periods, with $1 \leq T_0 < T$. Let Y_{it}^I be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods $T_0 + 1$ to T. We assume that the intervention has no effect on the outcome before the implementation period, so for $t \in \{1, \ldots, T_0\}$ and all $i \in \{1, \ldots, N\}$, we have that $Y_{it}^I = Y_{it}^N$.[5] Let $\alpha_{it} = Y_{it}^I - Y_{it}^N$ be the effect of the intervention for unit i at time t, if unit i is exposed to the intervention in periods $T_0+1, T_0+2, \ldots, T$ (where $1 \leq T_0 < T$). Therefore:

$$Y_{it}^I = Y_{it}^N + \alpha_{it}.$$

Let D_{it} be an indicator that takes value one if unit i is exposed to the intervention at time t, and value zero otherwise. The observed outcome for unit i at time t is

$$Y_{it} = Y_{it}^N + \alpha_{it} D_{it}.$$

Because only the first region (region "one") is exposed to the intervention and only after period T_0 (with $1 \leq T_0 < T$), we have that:

$$D_{it} = \begin{cases} 1 & \text{if } i = 1 \text{ and } t > T_0, \\ 0 & \text{otherwise.} \end{cases}$$

[5] Of course, this is done without loss of generality. If the anticipation of the intervention impacts the outcome before the intervention is implemented, we can always redefine T_0 to be the first period in which the outcome may possibly react to the (anticipated) intervention.

We aim to estimate $(\alpha_{1T_0+1}, \ldots, \alpha_{1T})$. For $t > T_0$,

$$\alpha_{1t} = Y^I_{1t} - Y^N_{1t} = Y_{1t} - Y^N_{1t}.$$

Because Y^I_{1t} is observed, to estimate α_{1t} we just need to estimate Y^N_{1t}. Suppose that Y^N_{it} is given by a factor model:

$$Y^N_{it} = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it}, \tag{2.1}$$

where δ_t is an unknown common factor with constant factor loadings across units, Z_i is a $(r \times 1)$ vector of observed covariates (not affected by the intervention), θ_t is a $(1 \times r)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is an $(F \times 1)$ vector of unknown factor loadings, and the error terms ε_{it} are unobserved transitory shocks at the region level with zero mean for all i.

It is important to notice that this model does not rule out the existence of time-varying measured determinants of Y^N_{it}. The vector Z_i may contain pre- and post-intervention values of time-varying variables, as long as they are not affected by the intervention. For example, suppose that $T = 2$, $T_0 = 1$, and that Z_{it} is a scalar random variable for $i = 1, \ldots, J+1$ and $t = 1, 2$. Then, if $Z_i = (Z_{i1} \;\, Z_{i2})'$, $\theta_1 = (\beta \;\, 0)$ and $\theta_2 = (0 \;\, \beta)$, we obtain $\theta_t Z_i = Z_{it}\beta$. Notice also that the model in (2.1) does not restrict Z_i, μ_i, and ε_{it} to be independent.

Consider a $(J \times 1)$ vector of weights $W = (w_2, \ldots, w_{J+1})'$ such that $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each particular value of the vector W represents a potential synthetic control, that is, a particular weighted average of control regions. The value of the outcome variable for each synthetic control indexed by W is:

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \delta_t + \theta_t \sum_{j=2}^{J+1} w_j Z_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}.$$

Suppose that there are $(w^*_2, \ldots, w^*_{J+1})$ (that sum to one) such that:

$$\sum_{j=2}^{J+1} w^*_j Y_{j1} = Y_{11}, \;\ldots\;, \sum_{j=2}^{J+1} w^*_j Y_{jT_0} = Y_{1T_0}, \;\text{ and }\; \sum_{j=2}^{J+1} w^*_j Z_j = Z_1. \tag{2.2}$$

Then, it is easy to see that if $\sum_{t=1}^{T_0} \lambda_t' \lambda_t$ is non-singular, then,

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt} = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \lambda_t \left(\sum_{n=1}^{T_0} \lambda_n' \lambda_n \right)^{-1} \lambda_s' (\varepsilon_{js} - \varepsilon_{1s}) - \sum_{j=2}^{J+1} w_j^* (\varepsilon_{jt} - \varepsilon_{1t}). \quad (2.3)$$

Appendix C shows that, under standard conditions, the average of the right hand side of equation (2.3) will be close to zero if the number of pre-intervention periods is large relative to the scale of the transitory shocks. This suggests using

$$\widehat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

for $t \in \{T_0 + 1, \ldots, T\}$ as an estimator of α_{1t}.

Equation (2.2) can hold exactly only if $(Y_{11}, \ldots, Y_{1T_0}, Z_1')$ belongs to the convex hull of $\{(Y_{21}, \ldots, Y_{2T_0}, Z_2'), \ldots, (Y_{J+11}, \ldots, Y_{J+1T_0}, Z_{J+1}')\}$. In practice, it is often the case that no set of weights exists such that equation (2.2) holds exactly in the data. Then, the synthetic control region is selected so that equation (2.2) holds approximately.

The simple linear model presented in this section does not need to hold over the entire set of regions in any particular sample. Researchers trying to minimize biases caused by interpolating across regions with very different characteristics may restrict the donor pool to regions with similar characteristics to the region exposed to the event or intervention of interest. As explained below, in contrast with more traditional regression methods, which typically rely on asymptotic limit theorems for inference, the availability of a small number of regions to construct the synthetic control does not invalidate our inferential procedures.

Notice that, even if taken at face value, equation (2.1) generalizes the usual difference-in-differences (fixed-effects) model commonly applied in the empirical literature. The difference-in-differences model allows for the presence of unobserved confounders but restricts the effect of those confounders to be constant in time. In contrast, the model presented in this section allows the effects of confounding unobserved characteristics to vary with time. Notice that the traditional difference-in-differences (fixed-effects) model can be obtained if we impose that λ_t in equation (2.1) is constant for all t.

Synthetic controls can provide useful estimates in more general contexts than the factor model considered thus far. Consider, for example, the following autoregressive model with time-varying coefficients:

$$\begin{aligned} Y^N_{it+1} &= \alpha_t Y^N_{it} + \beta_{t+1} Z_{it+1} + u_{it+1}, \\ Z_{it+1} &= \gamma_t Y^N_{it} + \Pi_t Z_{it} + v_{it+1}, \end{aligned} \tag{2.4}$$

where u_{it+1} and v_{it+1} have mean zero conditional on $\mathcal{F}_t = \{Y_{js}, Z_{js}\}_{1\leq j\leq N,\ s\leq t}$. Suppose that we can choose $\{w^*_j\}_{2\leq j\leq N}$ such that:

$$\sum_{j=2}^{J+1} w^*_j Y_{jT_0} = Y_{1T_0}, \quad \text{and} \quad \sum_{j=2}^{J+1} w^*_j Z_{jT_0} = Z_{1T_0}. \tag{2.5}$$

Then, it is easy to see that the synthetic control estimator is unbiased even if data for only a single pretreatment period are available.[6]

2.2.3 Implementation

Assume that there are J regions not exposed to the event or intervention of interest, so they can serve as controls. We consider any weighted average of non-exposed regions as a potential (synthetic) control. Let W be a $(J\times 1)$ vector of positive weights that sum to one. That is, $W = (w_2, \ldots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each value of W represents a weighted average of the available control regions and, therefore, a synthetic control.[7]

The outcome variable of interest is observed for T periods for the region affected by the intervention Y_{1t}, $(t = 1, \ldots, T)$ and the unaffected regions Y_{jt}, $(j = 2, \ldots, J+1, t = 1, \ldots, T)$. Let $T_1 = T - T_0$ be the number of post-intervention periods. Let Y_1 be the $(T_1 \times 1)$ vector of post-intervention outcomes for the exposed region, and Y_0 be the $(T_1 \times J)$ matrix of post-intervention outcomes for the potential control regions.

[6] See Appendix C for details.

[7] Although we define our synthetic controls as convex combinations of unexposed units, negative weights or weights larger than one can be used at the cost of allowing extrapolation. The severity of the extrapolation can be limited by specifying lower and upper bounds for the weights.

Let the $(T_0 \times 1)$ vector $K = (k_1, \ldots, k_{T_0})'$ define a linear combination of pre-intervention outcomes: $\bar{Y}_i^K = \sum_{s=1}^{T_0} k_s Y_{is}$.[8] Consider M of such linear combinations defined by the vectors $K_1, \ldots, K_M$. Let $X_1 = (Z_1', \bar{Y}_1^{K_1}, \ldots, \bar{Y}_1^{K_M})'$ be a $(k \times 1)$ vector of pre-intervention characteristics for the exposed region, with $k = r + M$. Similarly, X_0 is a $(k \times J)$ matrix that contains the same variables for the unaffected regions. That is, the j-th column of X_0 is $(Z_j', \bar{Y}_j^{K_1}, \ldots, \bar{Y}_j^{K_M})'$. The vector W^* is chosen to minimize some distance (or pseudo-distance), $\|X_1 - X_0 W\|$, between X_1 and $X_0 W$, subject to $w_2 \geq 0, \ldots, w_{J+1} \geq 0$, $w_2 + \cdots + w_{J+1} = 1$. In particular, we will consider $\|X_1 - X_0 W\|_V = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$, where V is some $(k \times k)$ symmetric and positive semidefinite matrix, although other choices are also possible.[9]

Although our inferential procedures are valid for any choice of V, the choice of V influences the mean square error of the estimator (that is, the expectation of $(Y_1 - Y_0 W^*)'(Y_1 - Y_0 W^*)$). Because V is symmetric and positive semidefinite, there exist two $(k \times k)$ matrices, U and A, such that the rows of U, $\{u_n\}_{n=1}^k$, form an orthonormal basis of $\mathbb{R}^k$, A is diagonal with all diagonal elements, $\{a_{nn}\}_{n=1}^k$, equal or greater than zero, and $V = U'AU$. As a result, the vector W^* minimizes $(H_1 - H_0 W)' A (H_1 - H_0 W)$, subject to $w_2 \geq 0, \ldots, w_{J+1} \geq 0$, $w_2 + \cdots + w_{J+1} = 1$, where $H_1 = UX_1$ and $H_0 = UX_0$. In other words, the matrix V assigns weight a_{nn} to the linear combination of characteristics in X_0 and X_1 with coefficients u_n. The optimal choice of V assigns weights to linear combination of the variables in X_0 and X_1 to minimize the mean square error of the synthetic control estimator. Sometimes this choice can be based on subjective assessments of the predictive power of the vari-

[8] For example, if $k_1 = k_2 = \cdots = k_{T_0 - 1} = 0$ and $k_{T_0} = 1$, then $\bar{Y}^K = Y_{iT_0}$, the value of the outcome variable in the period immediately prior to the intervention. If $k_1 = k_2 = \cdots = k_{T_0} = 1/T_0$, then $\bar{Y}_i^K = T_0^{-1} \sum_{s=1}^{T_0} Y_{is}$, the simple average of the outcome variable for the pre-intervention periods.

[9] If the relationship between the outcome variable and the explanatory variables in X_1 and X_0 is highly nonlinear and the support of the explanatory variables is large, interpolation biases may be severe. In that case, W^* can be chosen to minimize $\|X_1 - X_0 W\|$ plus a set of penalty terms specified as increasing functions of the distances between X_1 and the corresponding values for the control units with positive weights in W. Alternatively, as mentioned in section 2.2.2.2.2, interpolation biases can be reduced by restricting the comparison group to units that are similar to the exposed units in term of the values of X_1.

ables in X_1 and X_0. The choice of V can also be data-driven. One possibility is to choose V such that the resulting synthetic control region approximates the trajectory of the outcome variable of the affected region in the pre-intervention periods. For example, Abadie and Gardeazabal (2003) choose V among positive definite and diagonal matrices such that the mean squared prediction error of the outcome variable is minimized for the pre-intervention periods. Alternatively, if the number of available pre-intervention periods in the sample is large enough, researchers may divide them into an initial training period and a subsequent validation period. Given a V, $W^*(V)$ can be computed using data from the training period. Then, the matrix V can be chosen to minimize the mean squared prediction error produced by the weights $W^*(V)$ during the validation period.[10]

One obvious choice for $\bar{Y}_i^{K_1}, \ldots, \bar{Y}_i^{K_M}$ is $\bar{Y}_i^{K_1} = Y_{i1}, \ldots, \bar{Y}_i^{K_{T_0}} = Y_{iT_0}$. In practice, however, the computation of the weights $w_2^*, \ldots, w_{J+1}^*$ can be simplified by considering only a few linear combinations or pre-intervention outcomes and checking whether equation (2.2) holds approximately for those weights.

2.2.4 Inference

The standard errors commonly reported in regression-based comparative case studies measure uncertainty about aggregate data. For example, Card (1990) uses data from the U.S. Current Population Survey to estimate native employment rates in Miami and a set of comparison cities around the time of the Mariel Boatlift. Card and Krueger (1994) use data on a sample of fast-food restaurants in New Jersey and Pennsylvania to estimate the average number of employees in fast-food restaurants in these two states around the time when the minimum wage was increased in New Jersey. The standard errors reported in these studies reflect only the unavailability of aggregate data on employment (for native workers in Miami and other cities, and

[10] In cases when little is known about the relative predictive power of the pre-intervention variables, researchers may decide to normalize the variables in X_0 and X_1 using V equal to the inverse of the estimated variance-covariance matrix of the variables in X_0 and X_1 (Mahalanobis distance) or equal to a diagonal matrix with the elements in the main diagonal equal to the inverses of the sample variances of the variables in X_0 and X_1 (normalized Euclidean distance).

in fast-food restaurants in New Jersey and Pennsylvania, respectively). This mode of inference would logically produce zero standard errors if aggregate data were used for estimation. However, perfect knowledge of the value of aggregate data does not reduce to zero our uncertainty about the parameters of interest. That is, even if aggregate data are used for estimation, in most cases researchers would not believe that there is no remaining uncertainty about the value of the parameters of interest. The reason is that not all uncertainty about the value of the estimated parameters come from lack of knowledge of aggregate data. In comparative case studies, an additional source of uncertainty derives from our ignorance about the ability of the control group to reproduce the counterfactual of how the treated unit would have evolved in the absence of the treatment. This type of uncertainty is present regardless of whether aggregate data are used for estimation or not. The use of individual micro data, as opposed to aggregate data, only increases the amount of uncertainty if the outcome of interest is an aggregate.

Large sample inferential techniques are not well-suited to comparative case studies when the number of units in the comparison group and the number of periods in the sample are relatively small. In this article, we propose exact inferential techniques, akin to permutation tests, to perform inference in comparative case studies. The methods proposed here produce informative inference regardless of the number of available comparison units, the number of available time periods, and whether the data are individual (micro) or aggregate (macro). However, the quality of the inferential exercises proposed in this article increases with the number of available comparison units or the number of available time periods. The inferential techniques proposed in this article extend Abadie and Gardeazabal (2003) in several directions.

In their study of the economic effects of terrorism, Abadie and Gardeazabal (2003) use a synthetic control region to estimate the economic growth that the Basque Country would have experienced in the absence of terrorism. To assess the ability of the synthetic control method to reproduce the evolution of a counterfactual Basque Country without terrorism, Abadie and Gardeazabal (2003) introduce a placebo study, applying the same techniques to Catalonia, a region similar to the Basque Country but with a much lower exposure to terrorism. Similar falsification tests have been used to assess the effects of computers on the distribution of wages (DiNardo and Pischke, 1997), the effect of the Mariel Boatlift on native unemployment in Miami (Angrist and Krueger, 1999), and the validity of the rational addiction model for cigarette consumption (Auld and Grootendorst, 2004).[11]

In this paper, we extend the idea of a placebo study to produce quantitative inference in comparative case studies. As in classical permutation tests, we apply the synthetic control method to every potential control in our sample. This allows us to assess whether the effect estimated by the synthetic control for the region affected by the intervention is large relative to the effect estimated for a region chosen at random. This inferential exercise is exact in the sense that, regardless of the number of available comparison regions, time periods, and whether the data are individual or aggregate, it is always possible to calculate the exact distribution of the estimated effect of the placebo interventions. Notice also that the inferential exercise proposed here produces classical randomization inference for the case where the intervention is indeed randomized across regions, a rather restrictive condition. More generally, our inferential exercise examines whether or not the estimated effect of the actual intervention is large relative to the distribution of the effects estimated for the regions not exposed to the intervention. This is informative inference if under the hypothesis of no intervention effect the estimated effect of the intervention is not expected to be abnormal relative to the distribution of the placebo effects.[12] In this sense, our inferential procedure is related to that of DiNardo and Pischke (1997) and Auld and Grootendorst (2004). DiNardo and Pischke (1997) compare the wage differential associated with computer skills (as reflected in the on-the-job computer use) to the wage differentials associated with the use of other tools (pencils, telephones, calculators)

[11]What we refer to in this article as "placebo tests" or "falsification tests" sometimes appear with different names in the literature. Angrist and Krueger (1999) discuss empirical tests of this type under the heading "refutability" tests. Rosenbaum (2002a) discusses the use of outcomes "known to be unaffected by the treatment" to evaluate the presence of hidden biases.

[12]See Rosenbaum (2002a,b) for a detailed discussion of the use of permutation inference in randomized experiments and observational studies.

that do not proxy for skills that are scarce in the job market. Similarly, to assess the validity of the rational addiction model, Auld and Grootendorst (2004) compare the result of a test of rational addiction for cigarette consumption to the results of the same test applied to substances that are not considered addictive (milk, eggs, oranges, apples).

For cases in which the number of available comparison regions is very small small, one can use the longitudinal dimension of the data to produce placebo studies, as in Bertrand, Duflo, and Mullainathan (2004) where the dates of the placebo interventions are set at random.[13, 14]

2.3 Estimating the Effects of California's Proposition 99

2.3.1 Background

Anti-tobacco legislation has a long history in the United States, dating back at least as far as 1893, when Washington became the first state to ban the sale of cigarettes. Over the next 30 years 15 other states followed with similar anti-smoking measures (Dinan and Heckelman, 2005). These early anti-tobacco laws were primarily motivated by moral concerns; health issues were secondary (Tate, 1999). Almost 100 years later, after these early laws had long since been repealed, widespread awareness of smoking's health risks launched a new wave of state and federal anti-tobacco laws

[13]See also Heckman and Hotz (1989) for an earlier application of in-time placebos. In Appendix B we use an in-time placebo to study the economic impact of the 1990 German reunification.

[14]See Wooldridge (2003), Athey and Imbens (2006) and Donald and Lang (2007) for related work on inference in difference-in-differences models. Section 6.5 in Wooldridge and Imbens (2008) provides a recent survey of this literature. Conley and Taber (2008) propose an alternative method to do inference in comparative cases studies based on consistent estimation of the distribution of regression residuals for the case where the number of regions in the control group is large.

across the United States and, ultimately, overseas.[15] Leading this wave, in 1988, was a voter initiative in California known as Proposition 99, the first modern-time large-scale tobacco control program in the United States.

Proposition 99 increased California's cigarette excise tax by 25 cents per pack, earmarked the tax revenues to health and anti-smoking education budgets, funded anti-smoking media campaigns, and spurred local clean indoor-air ordinances throughout the state (Siegel, 2002).[16] Upon initial implementation, Proposition 99 produced more than $100 million per year in anti-tobacco projects for schools, communities, counties, and at the state level. Almost $20 million a year became available for tobacco-related research. As Glantz and Balbach (2000) put it, "[t]hese programs dwarfed anything that any other state or the federal government had ever done on tobacco."

Proposition 99 triggered a wave of local clean-air ordinances in California. Before Proposition 99 no city or town in California required restaurants to be 100 percent smoke-free. From 1989 to 2000 approximately 140 such laws were passed (Siegel, 2002). By 1993 local ordinances prohibiting smoking in the workplace protected nearly two-thirds of the workers in California (Glantz and Balbach, 2000). In 1994 the State of California passed additional legislation that banned smoking in enclosed workplaces. By 1996 more than 90 percent of California workers were covered by a smoke-free workplace policy (Siegel, 2002). Non-smokers' rights advocates view the wave of local ordinances passed under the impetus of Proposition 99 as an important step in the effort to undercut the then existing social support network for tobacco use in California (Glantz and Balbach, 2000).

The tobacco industry responded to Proposition 99 and the spread of clean-air ordinances by increasing its political activity in California at both the state and local levels. Tobacco lobby groups spent 10 times as much money in California in 1991-1992

[15]See Gruber (2001) for a survey on tobacco consumption and regulation in the United States. Ireland imposed a workplace smoking ban in 2004. This was followed by Italy in 2005, and Scotland in 2006. Belgium, Australia, and the United Kingdom have workplace smoking bans scheduled for 2007 (Borio, 2005).

[16]Proposition 99 assigned tax revenues to six accounts: Physician Services (35 percent), Health Education(20 percent), Hospital Services (10 percent), Research (5 percent), Public Resources (5 percent), and Unallocated (25 percent) (Glantz and Balbach 2000).

as they had spent in 1985-1986 (Begay et al., 1993). In addition, after the passage of Proposition 99, tobacco companies increased promotional expenditures in California (Siegel, 2002).

In 1991 California passed Assembly Bill 99, a new piece of legislation implementing Proposition 99. Contrary to the original mandate of Proposition 99, Assembly Bill 99 diverted a significant fraction of Health Education Account funds into medical services with little or no connection to tobacco (Glantz and Balbach, 2000). Also in 1991 a new governor began to exert increasing control over California's anti-smoking media campaign. In 1992 Governor Pete Wilson appointed a new Department of Health Services director and halted the media campaign, which provoked a lawsuit by the American Lung Association (ALA). The ALA won the suit and the campaign was back by the end of 1992, although with a reduced budget (Siegel, 2002).

Even so, Proposition 99 was widely perceived to have successfully cut smoking in California. From the passage of Proposition 99 through 1999 adult smoking prevalence fell in California by more than 30 percent, youth smoking levels dropped to the lowest in the country, and per capita cigarette consumption more than halved (California Department of Health Services, 2006). Prior to 1988 per capita cigarette consumption in California trailed the national average by 22.5 packs; ten years later per capita consumption was 40.4 packs lower than the national average (Siegel, 2002).

Following early reports of California's success with Proposition 99, other states adopted similar policies. In 1993 Massachusetts raised taxes on cigarettes from 26 to 51 cents per pack to fund a Health Protection Fund for smoking prevention and cessation programs. Similar laws passed in Arizona in 1995, with a 50-cent tax increase, and Oregon in 1997, where the tax on cigarettes rose from 38 to 68 cents per pack (Siegel, 2002). In November 1998 the tobacco companies signed a $206 billion Master Settlement Agreement that led the industry to impose an immediate 45-cent increase in cigarette prices nationwide (Capehart, 2001). As of October 6, 2006, there were 17 states and the District of Columbia and 519 municipalities across the country with laws in effect requiring 100 percent smoke-free workplaces, bars, or restaurants. Similar laws have been enacted but are not yet effective in other states

(ANRF, 2006).

Previous studies have investigated the impact of Proposition 99 on smoking prevalence using a variety of methods. Breslow and Johnson (1993), Glantz (1993), and Pierce et al. (1998) show that cigarette consumption in California after the passage of Proposition 99 in 1988 was lower than the average national trend and lower than the linearly extrapolated pre-program trend in California. Hu, Sung and Keeler (1995) use time-series regression to disaggregate the effects of Proposition 99's tax hike and media campaign on per-capita cigarette sales.

A related literature has studied the effect of smoking bans on smoking prevalence. Woodruff et al. (1993) show that smoking prevalence in California in 1990 was lower among workers affected by workplace smoking restrictions than among unaffected workers. More generally, Evans, Farrelly, and Montgomery, (1999), Farrelly, Evans, and Sfekas (1999), and Longo et al. (2001) have provided evidence on the effectiveness of workplace smoking bans.[17]

The most recently published study similar to ours is Fichtenberg and Glantz (2000), in the *New England Journal of Medicine.* This article uses least-squares regression to predict smoking rates in California as a function of the smoking rate for the rest of the United States. The regressions in Fichtenberg and Glantz (2000) estimate the effect of Proposition 99 as a time trend in per-capita cigarette consumption starting after the implementation of Proposition 99 in 1989. Fichtenberg and Glantz (2000) allow also for a change in this trend after 1992, when the anti-tobacco media campaign was first temporally eliminated and then reestablished but with reduced funds. Using this regression specification, Fichtenberg and Glantz (2000) estimate that during the period 1989-1992 Proposition 99 accelerated the rate of decline of per-capita cigarette consumption in California by 2.72 packs per year. Due to program cut-backs after 1992, Fichtenberg and Glantz (2000) estimate that during the period 1993-1997 Proposition 99 accelerated the rate of decline of per-capita cigarette consumption in California by only 0.67 packs per year.

[17]See also Goel and Nelson (2006) for a recent literature review on the effectiveness of anti-smoking legislation.

2.3.2 Data and Sample

We use annual state-level panel data for the period 1970-2000. Proposition 99 was passed in November 1988, giving us 18 years of pre-intervention data. Our sample period begins in 1970 because it is the first year for which data on cigarette sales are available for all our control states. It ends in 2000 because at about this time anti-tobacco measures were implemented across many states, invalidating them as potential control units. Moreover, a decade-long period after the passage of Proposition 99 seems like a reasonable limit on the span of plausible prediction of the effect of this intervention.

Recall that the synthetic California is constructed as a weighted average of potential control states, with weights chosen so that the resulting synthetic California best reproduces the values of a set of predictors of cigarette consumption in California before the passage of Proposition 99. Borrowing from the statistical matching literature, we refer to the set of potential controls for California as the "donor pool". Because the synthetic California is meant to reproduce the smoking rates that would have been observed for California in the absence of Proposition 99, we discard from the donor pool states that adopted some other large-scale tobacco control program during our sample period. Four states (Massachusetts, Arizona, Oregon, and Florida) introduced formal statewide tobacco control programs in the 1989-2000 period and they are excluded from the donor pool. We also discard all states that raised their state cigarette taxes by 50 cents or more over the 1989 to 2000 period (Alaska, Hawaii, Maryland, Michigan, New Jersey, New York, Washington).[18] Finally, we also exclude the District of Columbia from our sample. Our donor pool includes the remaining 38 states. Our results are robust, however, to the inclusion of discarded states.

Our outcome variable of interest is annual per capita cigarette consumption at the state level, measured in our dataset as per-capita cigarette sales in packs. We obtained these data from Orzechowski and Walker (2005) where they are constructed

[18]Notice that, even if the remaining tax increases substantially reduced smoking in any of the control states that gets assigned a positive W-weight, this should if anything attenuate the treatment effect estimate that we obtain for California.

using information on state-level tax revenues on cigarettes sales. This is the most widely used indicator in the tobacco research literature, available for a much longer time-period than survey-based measures of smoking prevalence. We include in X_1 and X_0 the values of predictors of smoking prevalence for California and the 38 potential controls, respectively. Our predictors of smoking prevalence are: average retail price of cigarettes, per-capita state personal income (logged), the percentage of the population age 15-24, and per-capita beer consumption. These variables are averaged over the 1980-1988 period, and augmented by adding three years of lagged smoking consumption (1975, 1980, and 1988). Appendix A provides data sources.[19, 20]

Using the techniques described in Section 2.2, we construct a synthetic California that mirrors the values of the predictors of cigarette consumption in California before the passage of Proposition 99. We estimate the effect of Proposition 99 on per-capita cigarette consumption as the difference in cigarette consumption levels between California and its synthetic versions in the years after Proposition 99 was passed. We then perform a series of placebo studies that confirm that our estimated effects for California are unusually large relative to the distribution of the estimate that we obtain when we apply the same analysis to all states in the donor pool.

2.3.3 Results

Figure 2.1 plots the trends in per-capita cigarette consumption in California and the rest of the United States. As this figure suggests, the rest of the United States may not provide a suitable comparison group for California to study the effects of Proposition 99 on per-capita smoking. Even before the passage of Proposition 99

[19]Average retail prices of cigarettes vary quite a bit across the United States. For example, in 1989, average retail prices ranged from \$1.16 in Kentucky to \$1.74 in Nevada.

[20]Results are robust regardless of which and how many predictor variables we include. The list of predictors used for robustness checks include: unemployment, income inequality, poverty, welfare transfers, crime rates, drug related arrest rates, state cigarette taxes, population density, and numerous variables to capture the demographic, racial, and social structure of states. Inclusion of these predictors leaves our results virtually unaffected. The weights associated with additional predictors in the matrix V usually are close to zero because the few predictors used in the streamlined baseline model already account for most of the variation in cigarette consumption over time.

the time series of cigarette consumption in California and in the rest of the United States differed notably. Levels of cigarette consumption were similar in California and the rest of the United States in the early 1970's. Trends began to diverge in the late 1970's, when California's cigarette consumption peaked and began to decline while consumption in the rest of the United States was still rising. Cigarette sales declined in the 1980's, but with larger decreases in California than in the rest of the United States. In 1988, the year Proposition 99 passed, cigarette consumption was about 27 percent higher in the rest of the United States relative to California. Following the law's passage cigarette consumption in California continued to decline. To evaluate the effect of Proposition 99 on cigarette smoking in California the central question is how cigarette consumption would have evolved in California after 1988 in the absence of Proposition 99. The synthetic control method provides a systematic way to estimate this counterfactual.

As explained above, we construct the synthetic California as the convex combination of states in the donor pool that most closely resembled California in terms of pre-Proposition 99 values of smoking prevalence predictors. The results are displayed in Table 2.1, which compares the pretreatment characteristics of the actual California with that of the synthetic California, as well as with the population-weighted average of the 38 states in the donor pool. We see that the average of states that did not implement a large-scale tobacco-control program in 1989-2000 does not seem to provide a suitable control group for California. In particular, prior to the passage of Proposition 99 average beer consumption and cigarette retail prices were lower in the 38 control states than in California. Moreover, prior to the passage of Proposition 99 average cigarette sales per-capita were substantially higher in the 38 control states than in California. In contrast, the synthetic California accurately reproduces the values that smoking prevalence and smoking prevalence predictor variables had in

California prior to the passage of Proposition 99.[21, 22]

Table 2.2 displays the weights of each control state in the synthetic California. The weights reported in Table 2.2 indicate that smoking trends in California prior to the passage of Proposition 99 is best reproduced by a combination of Colorado, Connecticut, Montana, Nevada, and Utah. All other states in the donor pool are assigned zero W-weights.

Figure 2.2 displays per-capita cigarette sales for California and its synthetic counterpart during the period 1970-2000. Notice that, in contrast to per capita sales in other U.S. states (shown in Figure 2.1), per-capita sales in the synthetic California very closely track the trajectory of this variable in California for the entire pre-Proposition 99 period. Combined with the high balance on all smoking predictors (Table 2.1), this suggests that the synthetic California provides a sensible approximation to the number of cigarette packs per-capita that would have been sold in California in 1989-2000 in the absence of Proposition 99.

Our estimate of the effect of Proposition 99 on cigarette consumption in California is the difference between per-capita cigarette sales in California and in its synthetic version after the passage of Proposition 99. Immediately after the law's passage, the two lines began to diverge noticeably. While cigarette consumption in the synthetic California continued on its moderate downward trend, the real California experienced a sharp downward kink. The discrepancy between the two lines suggests a large negative effect of Proposition 99 on per-capita cigarette sales. Figure 2.2 plots the yearly estimates of the impacts of Proposition 99, that is, the yearly gaps in per capita

[21] Table 2.1 highlights an attractive feature of synthetic control estimators. Similar to matching estimators, the synthetic control method forces the researcher to demonstrate the affinity between the region exposed to the intervention of interest and the regions in the donor pool. As a result, the synthetic control method safeguards against estimation of "extreme counterfactuals," that is, those counterfactuals that fall far outside the convex hull of the data (King and Zheng, 2006).

[22] We chose V among all positive definite and diagonal matrices to minimize the mean squared prediction error of per-capita cigarette sales in California during the pre-Proposition 99 period. The resulting value of the diagonal element of V associated to the log per-capita GDP variable is very small, which indicates that log GDP per-capita does not have substantial power predicting the per-capita cigarette consumption in California before the passage of Proposition 99. This explains the discrepancy between California and its synthetic version in terms of log GDP per-capita.

cigarette consumption between California and its synthetic counterpart. Figure 2.2 suggests that Proposition 99 had a large effect on per-capita cigarette sales, and that this effect increased in time. The magnitude of the estimated impact of Proposition 99 in Figure 2.2 is substantial. Our results suggest that for the entire 1989-2000 period cigarette consumption was reduced by an average of almost 20 packs per capita, a decline of approximately 25 percent.

Our analysis produces estimates of the effect of Proposition 99 that are considerably larger than those obtained by Fichtenberg and Glantz (2000) using linear regression methods. In particular, Fichtenberg and Glantz (2000) estimate that by 1997 Proposition 99 had reduced per-capita cigarette sales in California by about 14 packs per year. Our estimates increase this figure substantially, to 24 packs per year.[23]

2.3.4 Inference about the effect of the California Tobacco Control Program

To evaluate the statistical significance of our estimates, we pose the question of whether our results could be driven entirely by chance. How often would we obtain results of this magnitude if we had chosen a state at random for the study instead of California? To answer this question, we use placebo tests. Similar to Abadie and Gardeazabal (2003) and Bertrand, Duflo, and Mullainathan (2004), we run placebo studies by applying the synthetic control method to states that did not implement a large-scale tobacco control program during the sample period of our study. If the placebo studies create gaps of magnitude similar to the one estimated for California we interpret that our analysis does not provide significant evidence of a negative effect

[23]Part of this difference is likely to be explained by the fact that Fichtenberg and Glantz (2000) use per-capita cigarette sales in the rest of the United States to reproduce how this variable would have evolved in California in the absence of Proposition 99. As explained above, after the enactment of Proposition 99 in California, other states, like Massachusetts and Florida passed similar tobacco control legislation. While we eliminate these states as potential controls, Fichtenberg and Glantz (2000) do not do so, which is likely to attenuate their estimates.

of Proposition 99 on cigarette sales in California. If, on the other hand, the placebo studies show that the gap estimated for California is unusually large, relative to the gaps for the states that did not implement large-scale tobacco control program, we interpret that our analysis provides significant evidence of a negative effect of Proposition 99 on cigarette sales in California.

The idea of the placebo test is akin to the classic framework for permutation inference, where the distribution of a test statistic is computed under random permutations of the sample units' assignments to the intervention and non-intervention groups (see, for example, Lehmann, 1997).

To assess the significance of our estimates, we conduct a series of placebo studies by iteratively applying the synthetic control method used to estimate the effect of Proposition 99 in California to every other state in the donor pool. In each iteration we reassign in our data the tobacco control intervention to one of the 38 control states. That is, we proceed as if one of the states in the donor pool would have passed a large-scale tobacco control program in 1988, instead of California. We then compute the estimated effect associated with each placebo run. This iterative procedure provides us with a distribution of estimated gaps for the states in which no intervention took place.

Figure 2.4 displays the results for the placebo test. The gray lines represent the gap associated with each of the 38 runs of the test. That is, the gray lines show the difference in per-capita cigarette sales between each state in the donor pool and its respective synthetic version. The superimposed black line denotes the gap estimated for California. As the figure makes apparent, the estimated gap for California during the 1989-2000 period is unusually large relative to the distribution of the gaps for the states in the donor pool.

As Figure 2.4 indicates, the synthetic method provides an excellent fit for per-capita cigarette sales in California prior to the passage of Proposition 99. The pre-intervention mean squared prediction error (MSPE) in California (the average of the squared discrepancies between per-capita cigarette sales in California and in its synthetic counterpart during the period 1970-1988) is about 3. The pre-Proposition

99 median MSPE among the 38 states in the donor pool is about 6, also quite small, indicating that the synthetic control method is able to provide a good fit for per-capita cigarette consumption prior to Proposition 99 for the majority of the states in the donor pool. However, Figure 2.4 indicates also that per-capita cigarette sales during the 1970-1988 period cannot be well-reproduced for some states by a convex combination of per-capita cigarette sales in other states. The state with worst fit in the pre-Proposition 99 period is New Hampshire, with a MSPE of 3,437. The large MSPE for New Hampshire does not come as a surprise. Among all the states in the donor pool, New Hampshire is the state with the highest per-capita cigarette sales for every year prior to the passage of Proposition 99. Therefore, there is no combination of states in our sample that can reproduce the time series of per-capita cigarette sales in New Hampshire prior to 1988. Similar problems arise for other states with extreme values of per-capita cigarette sales during the pre-Proposition 99 period.

If the synthetic California had failed to fit per-capita cigarette sales for the real California in the years before the passage of Proposition 99 we would have interpreted that much of the post-1988 gap between the real and the synthetic California was also artificially created by lack of fit, rather than by the effect of Proposition 99. Similarly, placebo runs with poor fit prior to the passage of Proposition 99 do not provide information to measure the relative rarity of estimating a large post-Proposition 99 gap for a state that was well-fitted prior to Proposition 99. For this reason, we provide several different versions of Figure 2.4, each version excluding states beyond a certain level of pre-Proposition 99 MSPE.

Figure 2.5 excludes states that had a pre-Proposition 99 MSPE of more than 20 times the MSPE of California. This is a very lenient cutoff, discarding only four states with extreme values of pre-Proposition 99 MSPE, and for which the synthetic method would be clearly ill-advised. In this figure there remain a few lines that still deviate substantially from the zero gap line in the pre-Proposition 99 period. Among the 35 states remaining in the figure, the California gap line is now about the most unusual line, especially from the mid 1990's onward.

Figure 2.6 is based on a lower cutoff, excluding all states that had a pre-Proposition

99 MSPE of more than five times the MSPE of California. Twenty-nine control states plus California remain in the figure. The California gap line is now clearly the most unusual line for almost the entire post-treatment period.

In Figure 2.7 we lower the cutoff even further and focus exclusively on those states that we can fit almost as well as California in the period 1970-1988, that is, those states with pre-Proposition 99 MSPE not higher than twice the pre-Proposition 99 MSPE for California. Evaluated against the distribution of the gaps for the 19 remaining control states in Figure 2.7, the gap for California appears highly unusual. The negative effect in California is now by far the lowest of all. Because this figure includes 19 control states, the probability of estimating a gap of the magnitude of the gap for California under a random permutation of the intervention in our data is 5 percent, a test level typically used in conventional tests of statistical significance.

One final way to evaluate the California gap relative to the gaps obtained from the placebo runs is to look at the distribution of the ratios of post/pre-Proposition 99 MSPE. The main advantage of looking at ratios is that it obviates choosing a cutoff for the exclusion of ill-fitting placebo runs. Figure 2.8 displays the distribution of the post/pre-Proposition 99 ratios of the MSPE for California and all 38 control states. The ratio for California clearly stands out in the figure: post-Proposition 99 MSPE is about 130 times the MSPE for the pre-Proposition 99 period. No control state achieves such a large ratio. If one were to assign the intervention at random in the data, the probability of obtaining a post/pre-Proposition 99 MSPE ratio as large as California's is $1/39 = 0.026$.

2.4 Conclusion

Comparative case study research has broad potential in the social sciences. However, the empirical implementation of comparative case studies is plagued by inferential challenges and ambiguity about the choice of valid control groups. In this paper, we propose data-driven procedures to select synthetic comparison units in comparative case studies. We show that the synthetic control estimator is valid under fairly general conditions. In addition, we propose a method to produce inference in comparative cases studies that incorporates uncertainty about the validity of the control unit. Moreover, we provide software to implement the estimators proposed in this article.

We demonstrate the applicability of the synthetic control method by studying the effects of Proposition 99, a large-scale tobacco control program that California passed in 1988. Our results suggest the effects of the tobacco control program are much larger than prior estimates have reported. We show that if one were to re-label the intervention state in the dataset at random, the probability of obtaining results of the magnitude of those obtained for California would be extremely small, 0.026.

2.5 Appendix A: Data Sources

In this appendix, we describe the data used in our analysis and provide sources.

- Per-capita cigarette consumption (in packs). Source: Orzechowski and Walker (2005). These data are based on the total tax paid on sales of packs of cigarettes in a particular state divided by its total population.
- Average retail price per pack of cigarettes. Source: Orzechowski and Walker (2005). Price figures include state sales taxes, if applicable.
- Per-capita state personal income (logged). Source: Bureau of the Census, United States Statistical Abstract. Converted to 1997 dollars using the Consumer Price Index.
- State population and percent of state population aged 15-24. Source: U.S. Census Bureau.
- Per-capita beer consumption. Source: Beer Institute's Brewer's Almanac. Measured as the per-capita consumption of malt beverages (in gallons).

2.6 Appendix B: The Economic Impact of the German Reunification in West Germany

In this appendix, we illustrate the application of the synthetic control method to cross-country data. For this purpose, we apply the synthetic control method to estimate the economic impact of the 1990 German reunification in the former West Germany (Federal Republic of Germany). Using the synthetic control method, we construct a synthetic West Germany as a convex combination of other advanced industrialized countries chosen to resemble the values of economic growth predictors for West Germany prior to the reunification. Our sample of potential controls includes the following OECD member countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Greece, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and the United States.

We provide a list of all variables used in the analysis at the end of this appendix, along with data sources. For each variable we have made sure that the German data refers exclusively to the territory of the former West Germany. For that purpose, when necessary, our dataset was supplemented with data from the German Federal Statistical Office (*Statistisches Bundesamt*). The outcome variable is real per-capita GDP (PPP-adjusted, measured in 2002 U.S. Dollars). We rely on a standard set of predictors commonly used in the economic growth literature. We use the investment rate measured as the ratio of real domestic investment (private plus public) to GDP. Our proxy for human capital is the percentage of high school graduates in the total population aged 25 and older. Our predictors also include the inflation rate and the share of value added by the industrial sector. In addition, we include trade openness, measured by the sum of exports and imports as a percentage of GDP, among our predictors of economic growth. The inclusion of additional growth predictors did not change our results substantively.

Table 2.3 compares the pre-reunification characteristics of the actual West Germany to those of our synthetic West Germany, and also to those of the population-weighted average of the 20 OECD countries in the donor pool. The statistics in this table show that the synthetic West Germany approximates the pre-1990 values of the economic growth predictors for West Germany far more accurately than the average of our sample of other OECD countries. This suggests that the synthetic West Germany provides a better control for the actual West Germany than the average of our sample of other OECD countries.

Table 2.4 shows the weights of each country in the synthetic West Germany. The synthetic West Germany is a weighted average of Austria, the United States, Switzerland, the Netherlands, and Japan, with weights decreasing in this order. All other countries in the donor pool are assigned zero weights.

Figure 2.9 displays the GDP per-capita trajectory of West Germany and its synthetic counterpart for the 1960-2003 period. The synthetic West Germany almost exactly reproduces the trend of this variable in the actual West Germany for the entire pre-reunification period. This remarkable fit for the pre-treatment period, along with the high balance on GDP predictors (Table A.1), suggests that our synthetic West Germany provides a sensible estimate of the counterfactual GDP per-capita trend that West Germany would have experienced in the absence of the German reunification.

Our estimate of the effect of the German reunification on GDP per-capita in West Germany is given by the difference between the actual West Germany and its synthetic version. We find that the German reunification did not have much of an effect on West German GDP per-capita in the first two years immediately following reunification. In this initial period GDP per-capita in the synthetic West Germany is slightly lower than in the actual West Germany, which is broadly in line with arguments about an initial demand boom (see, for example, Meinhardt et al. 1995). From 1992 onwards, however, the two lines diverge substantially. While the actual West Germany's per-capita GDP growth decelerates, the synthetic West Germany's per-capita GDP keeps ascending at a pace similar to that of the pre-unification period. The divergence of the two series continues to grow until the end of the sample period. Our results suggest a pronounced negative effect of the reunification on West German income.

They suggest that for the entire 1990-2003 period, GDP per-capita was reduced by about $1,460 per year on average, a decline of approximately 6 percent.

To evaluate the statistical significance of our estimate, we conduct a placebo study where the treatment is reassigned not across units but in time. Thus, instead of running a placebo study by reassigning reunification to different countries, we maintain West Germany as the treated unit but reassign reunification to an earlier point in time. The idea is to evaluate how likely it is to obtain results of the magnitude that we obtain in Figure A.1 when we apply our methods in a sample period before the reunification.

Here we show results for the case when reunification is reassigned to the year 1980, ten years earlier than it actually occurred. Thus, we run the same model as before, but the pre-treatment period is constrained to be 1960-1980. We also lag our predictors variables accordingly. The results for reassigning reunification to years other than 1980 are substantially identical. They are omitted to economize on space.

Figure 2.10 displays the results of our in-time placebo study. Notice that the synthetic West Germany almost exactly reproduces the evolution of GDP per capita in the actual West Germany for the entire pre-treatment period. Most importantly, the GDP per-capita trajectories of West Germany and its synthetic counterpart do not diverge considerably after 1980. That is, in contrast to the actual 1990 German reunification, our 1980 placebo reunification has no perceivable effect. This placebo study shows that a weighted combination of other OECD countries can be used to predict accurately the evolution of GDP per-capita in West Germany prior to the reunification. This suggests that the gap estimated in Figure 2.9 reflects the impact of the German reunification and not a potential lack of predictive power of the synthetic control.

The data sources employed for this application are:

- GDP per capita (PPP 2002 $US). Source: OECD National Accounts (retrieved via the OECD Health Database in October 2006). Data for West Germany was obtained from Statistisches Bundesamt 2005 (Arbeitskreis "Volkswirtschaftliche Gesamtrechnungen der Länder") and converted using PPP monetary conversion factors (retrieved from the OECD Health Database in October 2006).

- Investment Rate: Ratio of real domestic investment (private plus public) to real GDP. Source: Barro and Lee (1994).

- Schooling: Percentage of secondary school completed in the total population aged 25 and older. Source: Barro and Lee (2000).

- Industry: industry share of value added. Source: World Bank WDI Database 2005 and Statistisches Bundesamt 2005.

- Inflation: annual percentage change in consumer prices (base year 1995). Source: World Development Indicators Database 2005 and Statistisches Bundesamt 2005.

- Trade Openness: Export plus Imports as percentage of GDP. Source: World Bank: World Development Indicators CD-ROM 2000.

2.7 Appendix C: Technical Details

Consider the first model in section 2.2.2.2.2,

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it},$$

where $\lambda_t = (\lambda_{t1}, \ldots, \lambda_{tF})$ is a $(1 \times F)$ vector of common factors, for $t = 1, \ldots, T$, and $\mu_i = (\mu_{i1}, \ldots, \mu_{iF})'$ is an $(F \times 1)$ vector of factor loadings, for $i = 1, \ldots, J+1$. The weighted average of the outcome in the donor pool, using weights $\{w_j\}_{2 \leq j \leq J+1}$ is:

$$\sum_{j=2}^{J+1} w_j Y_{jt}^N = \delta_t + \theta_t \left(\sum_{j=2}^{J+1} w_j Z_j \right) + \lambda_t \left(\sum_{j=2}^{J+1} w_j \mu_j \right) + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}.$$

As a result,

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}^N = \theta_t \left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j \right) + \lambda_t \left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j \right) + \sum_{j=2}^{J+1} w_j (\varepsilon_{1t} - \varepsilon_{jt}).$$

To simplify the exposition, we will assume that the terms ε_{it} are independent across units and in time. The analysis can, however, be extended to more general settings. Notice that even with ε_{it} independent across units and in time, the unobserved residual $u_{it} = \lambda_t \mu_i + \varepsilon_{it}$ may be correlated across units and in time because the presence of the term $\lambda_t \mu_i$. Assume also that the terms ε_{it} are mean-independent of $\{Z_i, \mu_i\}_{i=1}^{J+1}$. Let Y_i^P be the $T_0 \times 1$ vector with t-th element equal to Y_{it}. Similarly, let ε_i^P be the $(T_0 \times 1)$ vector with t-th element equal to ε_{it}. Finally, let θ^P and λ^P be the $(T_0 \times r)$ matrix and $(T_0 \times F)$ matrix with t-th rows equal to θ_t and λ_t, respectively. We obtain,

$$Y_1^P - \sum_{j=2}^{J+1} w_j Y_j^P = \theta^P \left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j \right) + \lambda^P \left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j \right) + \sum_{j=2}^{J+1} w_j (\varepsilon_1^P - \varepsilon_j^P).$$

Let $\xi(M)$ be the smallest eigenvalue of:

$$\frac{1}{M} \sum_{t=T_0-M+1}^{T_0} \lambda_t' \lambda_t.$$

Assume that $\xi(M)$ is bounded away from zero: $\xi(M) \geq \underline{\xi} > 0$, for each positive integer, M. Assume also that $|\lambda_{tf}| \leq \bar{\lambda}$ for all $t = 1, \ldots, T$, $f = 1, \ldots, F$. Therefore, because $\lambda^{P\prime}\lambda^P$ is not singular:

$$\begin{aligned} Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}^N &= \lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\left(Y_1^P - \sum_{j=2}^{J+1} w_j Y_j^P\right) \\ &+ (\theta_t - \lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\theta^P)\left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j\right) \\ &- \lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\left(\varepsilon_1^P - \sum_{j=2}^{J+1} w_j \varepsilon_j^P\right) + \sum_{j=2}^{J+1} w_j(\varepsilon_{1t} - \varepsilon_{jt}). \end{aligned}$$

Suppose that there exist $\{w_2^*, \ldots, w_{J+1}^*\}$ such that equation (2.2) holds. Then

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N = R_{1t} + R_{2t} + R_{3t},$$

where

$$R_{1t} = \lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\sum_{j=2}^{J+1} w_j^* \varepsilon_j^P, \quad R_{2t} = -\lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\varepsilon_1^P,$$

and $R_{3t} = \sum_{j=2}^{J+1} w_j^*(\varepsilon_{jt} - \varepsilon_{1t})$. Consider the case of $t > T_0$. Then, R_{2t} and R_{3t} have mean zero. Notice that,

$$R_{1t} = \sum_{j=2}^{J+1} w_j \sum_{s=1}^{T_0} \lambda_t \left(\sum_{n=1}^{T_0} \lambda_n' \lambda_n\right)^{-1} \lambda_s' \varepsilon_{js}.$$

Because $\sum_{t=1}^{T_0} \lambda_t'\lambda_t$ is symmetric and positive definite, so is its inverse. Then, applying the Cauchy-Schwarz Inequality, we obtain:

$$\begin{aligned} \left(\lambda_t \left(\sum_{n=1}^{T_0} \lambda_n'\lambda_n\right)^{-1} \lambda_s'\right)^2 &\leq \left(\lambda_t \left(\sum_{n=1}^{T_0} \lambda_n'\lambda_n\right)^{-1} \lambda_t'\right)\left(\lambda_s \left(\sum_{n=1}^{T_0} \lambda_n'\lambda_n\right)^{-1} \lambda_s'\right) \\ &\leq \left(\frac{\bar{\lambda}^2 F}{T_0 \underline{\xi}}\right)^2. \end{aligned}$$

Let

$$\bar{\varepsilon}_j^L = \sum_{s=1}^{T_0} \lambda_t \left(\sum_{n=1}^{T_0} \lambda_n'\lambda_n\right)^{-1} \lambda_s' \varepsilon_{js}$$

for $j = 2, \ldots, J+1$.

Assume that, for some even p, the p-th moments of $|\varepsilon_{jt}|$ exist for $j = 2, \ldots, J+1$ and $t = 1, \ldots, T_0$. Using Hölder's Inequality:

$$\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L| \leq \left(\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L|^p \right)^{1/p} \leq \left(\sum_{j=2}^{J+1} |\bar{\varepsilon}_j^L|^p \right)^{1/p}.$$

Therefore, applying again Hölder's Inequality:

$$E\left[\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L| \right] \leq \left(E\left[\sum_{j=2}^{J+1} |\bar{\varepsilon}_j^L|^p \right] \right)^{1/p}.$$

Now, using Rosenthal's Inequality:

$$E|\bar{\varepsilon}_j^L|^p \leq C(p) \left(\frac{\bar{\lambda}^2 F}{\underline{\xi}} \right)^p \max \left\{ \frac{1}{T_0^p} \sum_{t=1}^{T_0} E|\varepsilon_{jt}|^p, \left(\frac{1}{T_0^2} \sum_{t=1}^{T_0} E|\varepsilon_{jt}|^2 \right)^{p/2} \right\},$$

where $C(p)$ is the p-th moment of minus one plus a Poisson random variable with parameter equal to one (see Ibragimov and Sharakhmetov, 2002). Let $\sigma_{jt}^2 = E|\varepsilon_{jt}|^2$, $\sigma_j^2 = (1/T_0) \sum_{t=1}^{T_0} \sigma_{jt}^2$, $\bar{\sigma}^2 = \max_{j=2,\ldots,J+1} \sigma_j^2$, and $\bar{\sigma} = \sqrt{\bar{\sigma}^2}$. Similarly, let $m_{p,jt} = E|\varepsilon_{jt}|^p$, $m_{p,j} = (1/T_0) \sum_{t=1}^{T_0} m_{p,jt}$, and $\bar{m}_p = \max_{j=2,\ldots,J+1} m_{p,j}$. We obtain that, for $t > T_0$,

$$E|R_{1t}| \leq C(p)^{1/p} \left(\frac{\bar{\lambda}^2 F}{\underline{\xi}} \right) J^{1/p} \max \left\{ \frac{\bar{m}_p^{1/p}}{T_0^{1-1/p}}, \frac{\bar{\sigma}}{T_0^{1/2}} \right\}.$$

Last equation shows that the bias of the estimator can be bounded by a function that goes to zero as the number of pre-treatment periods increases.

Consider now the autoregressive model in equation (2.4). Notice that

$$Y_{iT_0+1}^N = (\alpha_{T_0} + \beta_{T_0+1} \gamma_{T_0}) Y_{iT_0} + \beta_{T_0+1} \Pi_{T_0} Z_{iT_0} + \beta_{T_0+1} v_{iT_0+1} + u_{iT_0+1},$$

where $\{u_{it}, v_{it}\}_{T_0+1 \leq t \leq T_0+n}$ have mean zero conditional on $\mathcal{F}_{T_0}$. Working recursively it is easy to check that, conditional on Y_{iT_0} and Z_{iT_0}, for $n \geq 1$, $Y_{iT_0+n}^N$ is a linear function of $\{u_{it}, v_{it}\}_{T_0+1 \leq t \leq T_0+n}$. Then, because $\{w_j^*\}_{2 \leq j \leq N}$ is a deterministic function of $\mathcal{F}_{T_0}$, the bias of the synthetic control estimator goes to zero as the discrepancies in equation in equation (2.5) go to zero.

2.8 Tables for Chapter 2

Table 2.1: Cigarette Sales Predictor Means

	California		Average of
Variables	Real	Synthetic	38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15-24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales Per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

Note: All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988).

Table 2.2: State Weights in the Synthetic California

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	-	Nebraska	0
Arizona	-	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	-
Connecticut	0.069	New Mexico	0
Delaware	0	New York	-
District of Columbia	-	North Carolina	0
Florida	-	North Dakota	0
Georgia	0	Ohio	0
Hawaii	-	Oklahoma	0
Idaho	0	Oregon	-
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	-	Vermont	0
Massachusetts	-	Virginia	0
Michigan	-	Washington	-
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

Table 2.3: Economic Growth Predictor Means before the German Reunification

	West Germany	Synthetic West Germany	OECD Sample excl. West Germany
GDP per-capita	10774.85	10808.02	10161.86
Inflation rate	3.80	5.17	9.01
Trade openness	50.59	58.41	33.47
Schooling	29.20	31.21	22.37
Investment rate	27.00	27.00	25.44
Industry share	34.69	34.64	34.28

Note: GDP, inflation rate, and trade openness are averaged for the 1970–1989 period. Industry share is averaged for the 1980–1989 period. Investment rate is averaged for the 1980–1985 period. Schooling is from year 1985.

Table 2.4: Country Weights in the Synthetic West Germany

Country	Weight	Country	Weight
Australia	0	Japan	0.127
Austria	0.421	Netherlands	0.137
Belgium	0	New Zealand	0
Canada	0	Norway	0
Denmark	0	Portugal	0
Finland	0	Spain	0
France	0	Sweden	0
Greece	0	Switzerland	0.153
Ireland	0	United Kingdom	0
Italy	0	United States	0.161

2.9 Figures for Chapter 2

Figure 2.1: Trends in Per-Capita Cigarette Sales: California vs the Rest of the United States

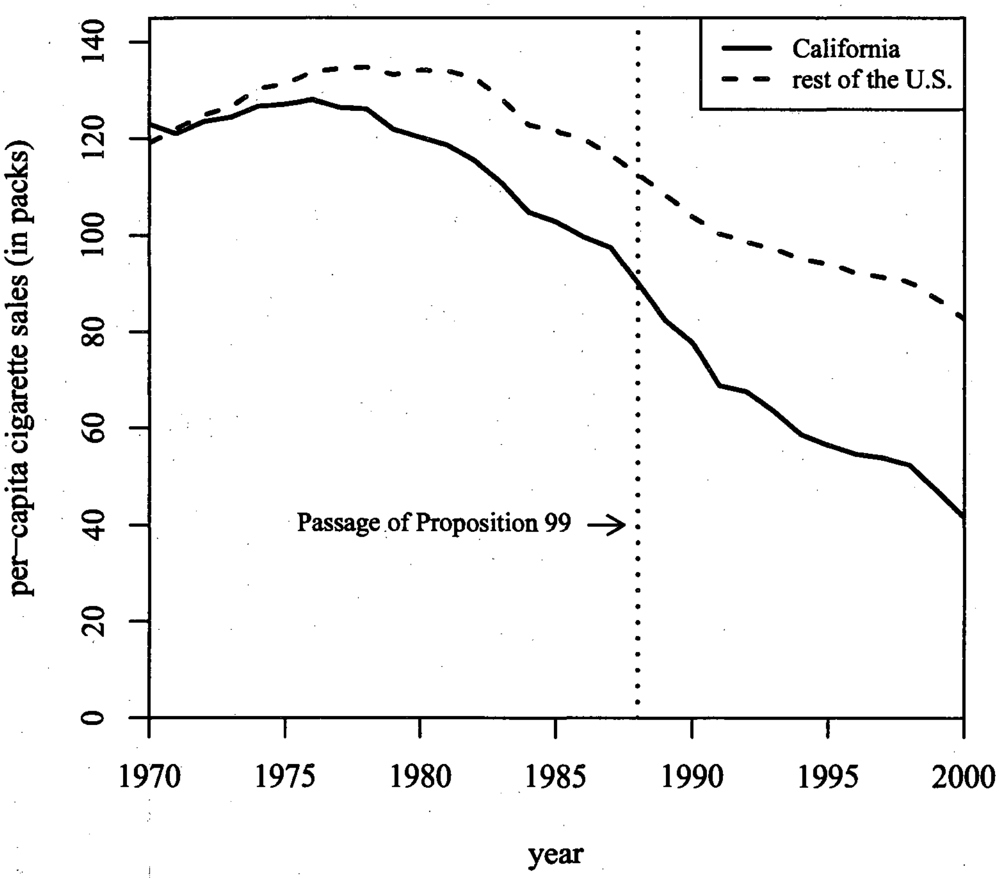


Figure 2.2: Trends in Per-Capita Cigarette Sales: California vs. synthetic California

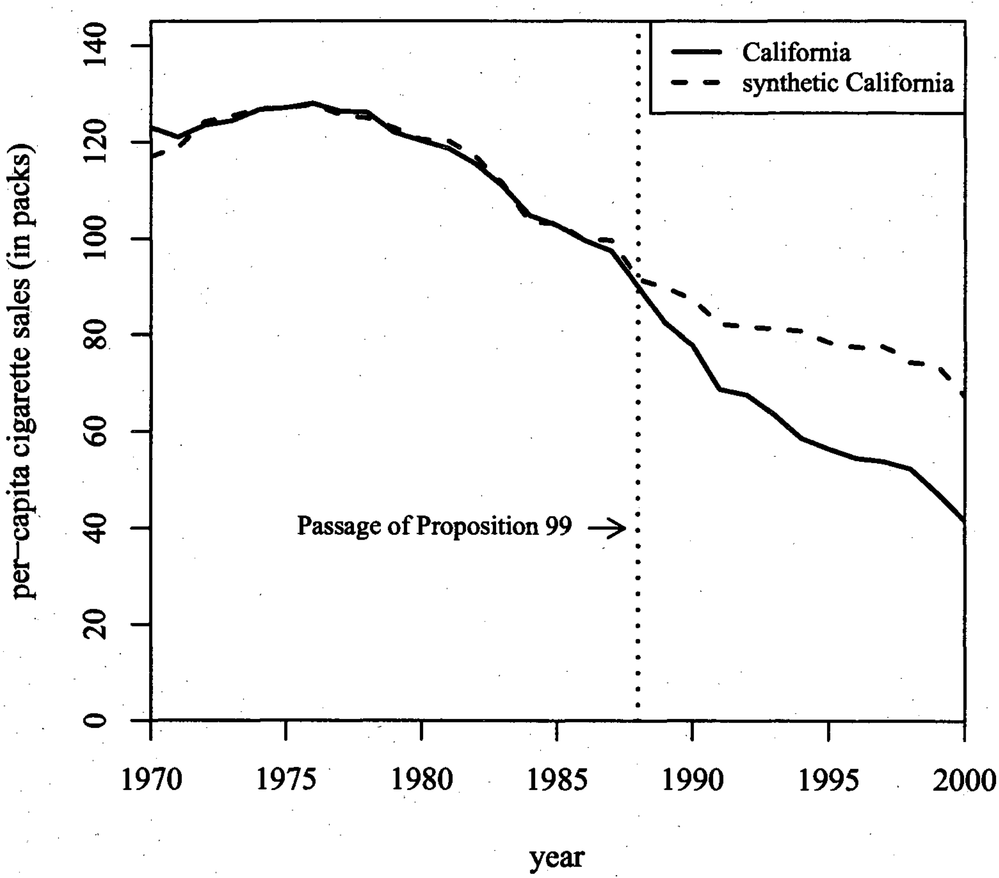


Figure 2.3: Per-Capita Cigarette Sales Gap Between California and Synthetic California

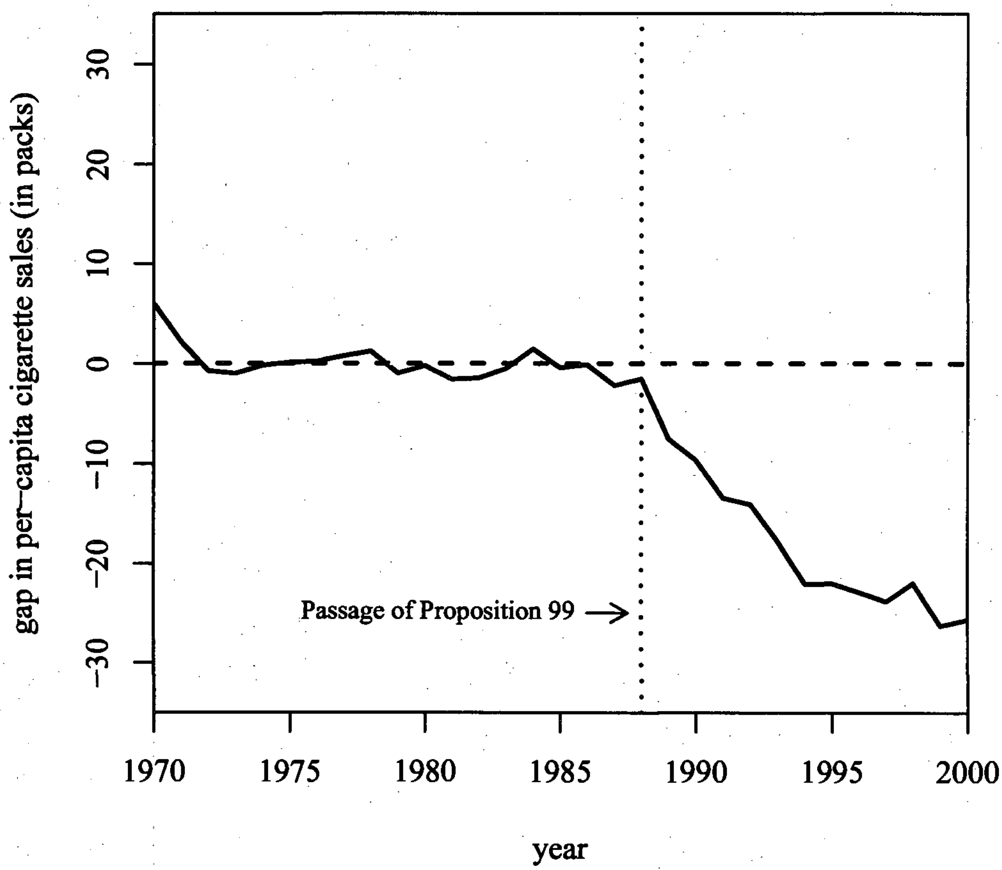


Figure 2.4: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in all 38 Control States

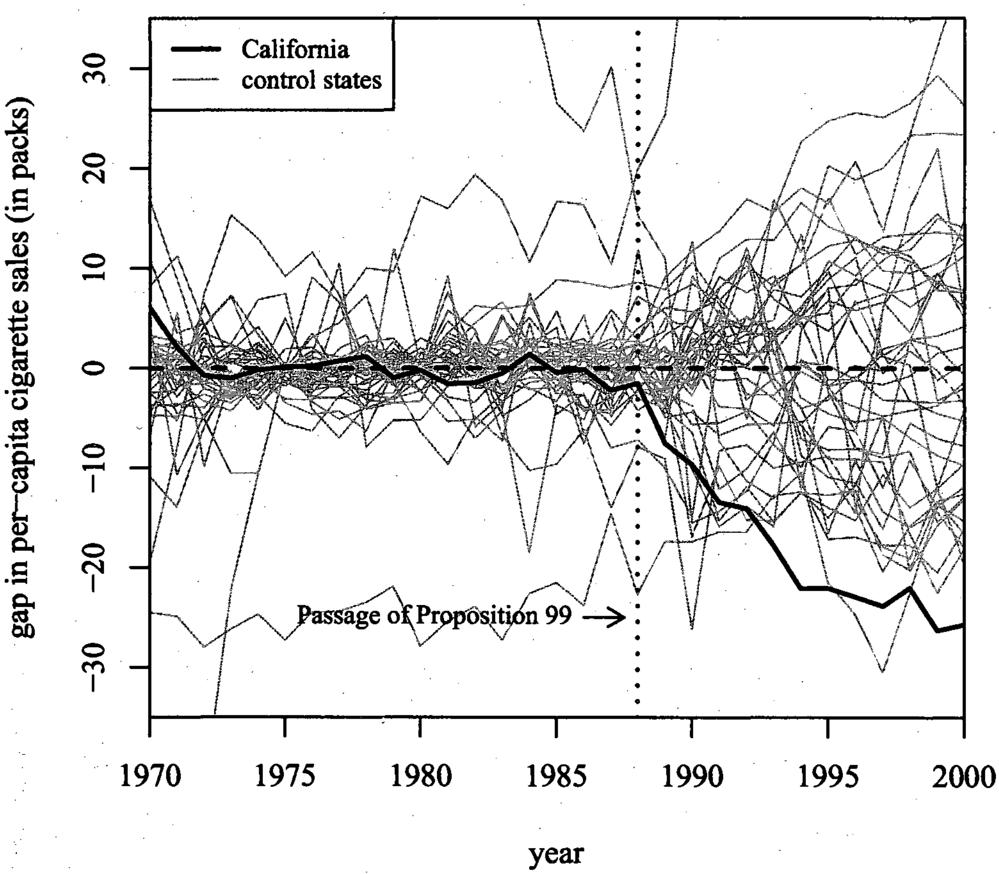


Figure 2.5: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 34 Control States (Discards States with Pre-Proposition 99 MSPE Twenty Times Higher than California's)

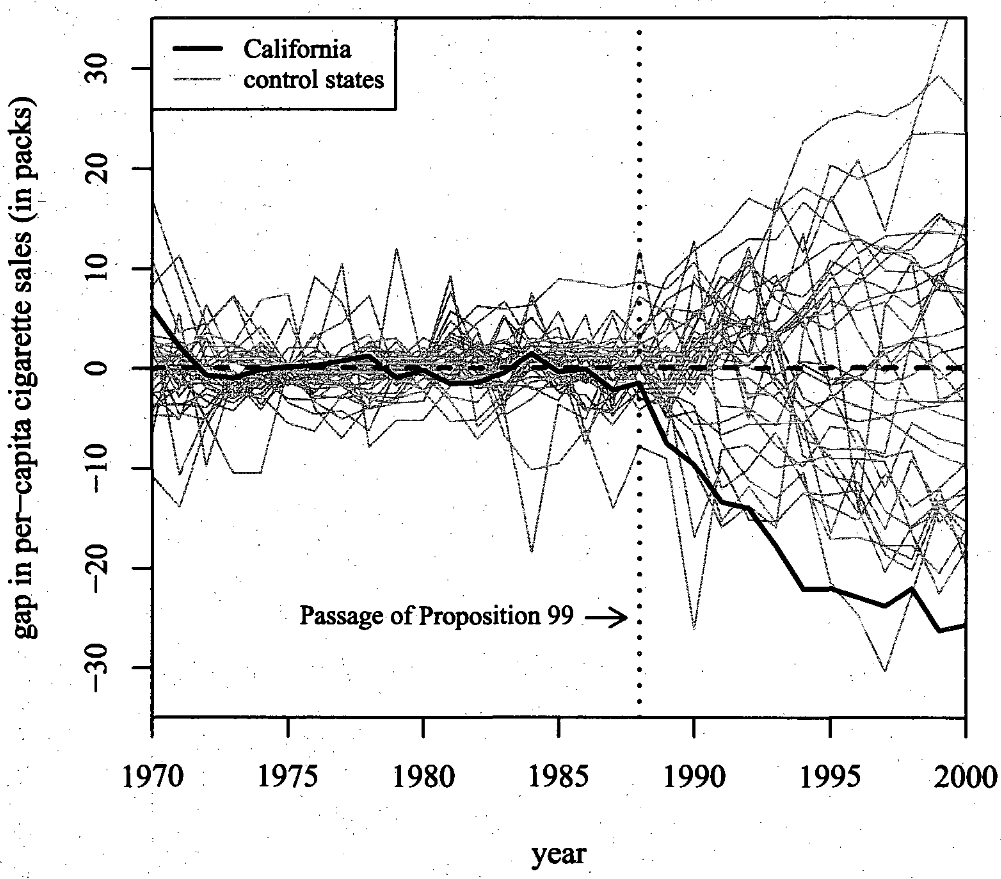


Figure 2.6: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 29 Control States (Discards States with Pre-Proposition 99 MSPE Five Times Higher than California's)

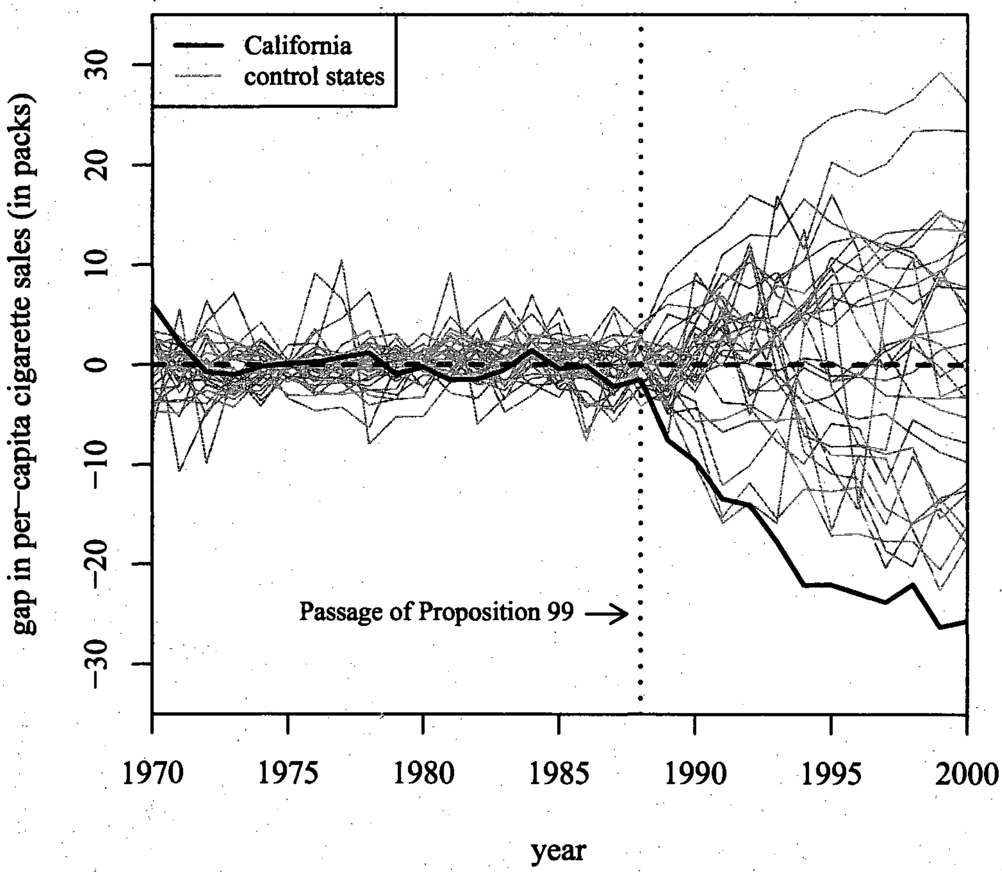


Figure 2.7: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 19 Control States (Discards States with Pre-Proposition 99 MSPE Two Times Higher than California's)

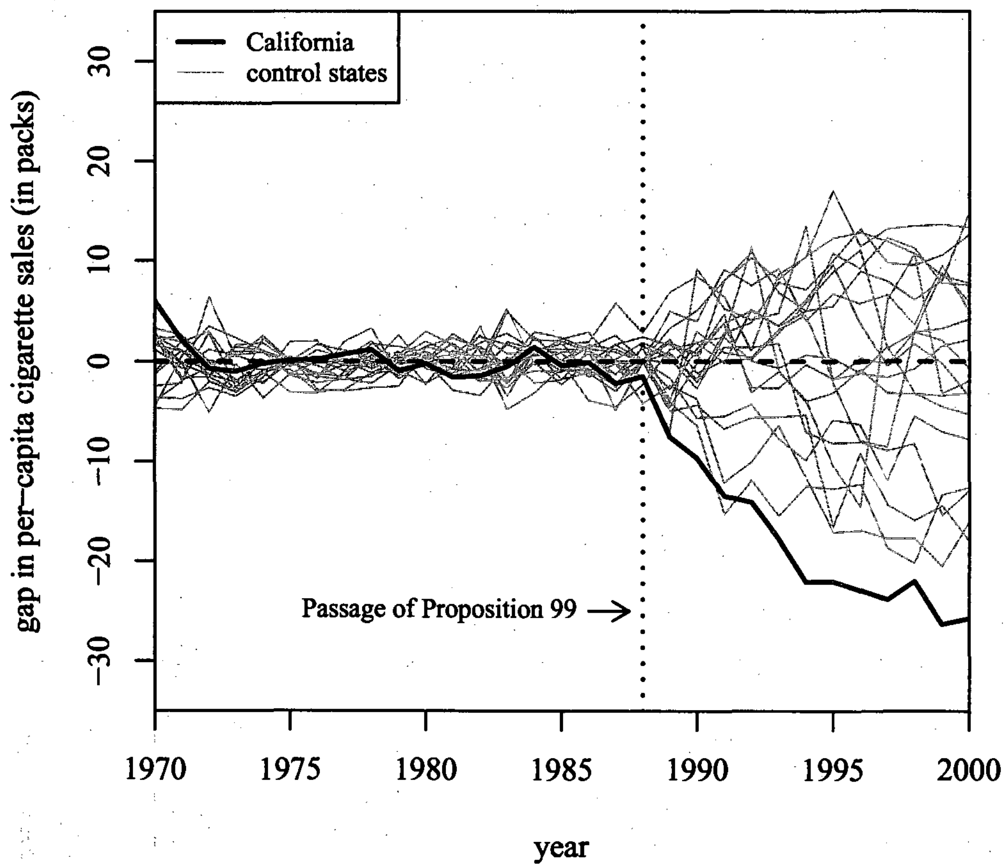


Figure 2.8: Ratio of Post-Proposition 99 MSPE and Pre-Proposition 99 MSPE: California and 38 Control States

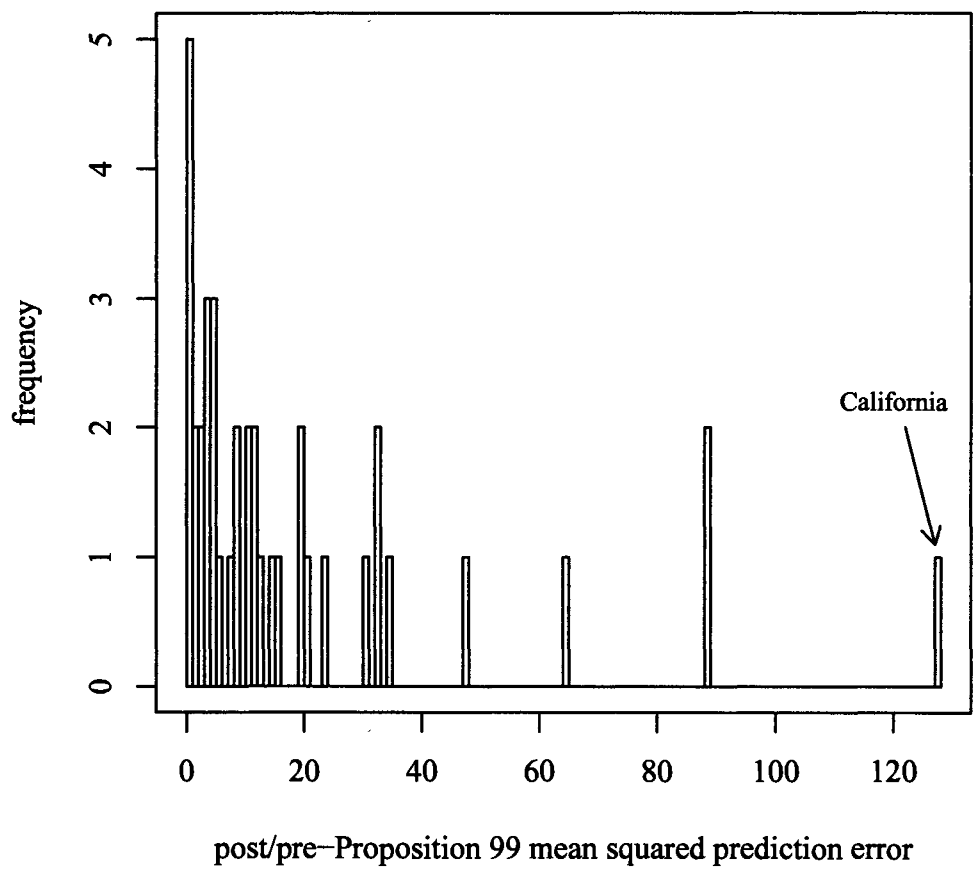


Figure 2.9: Trends in Per-Capita GDP: West Germany vs. Synthetic West Germany

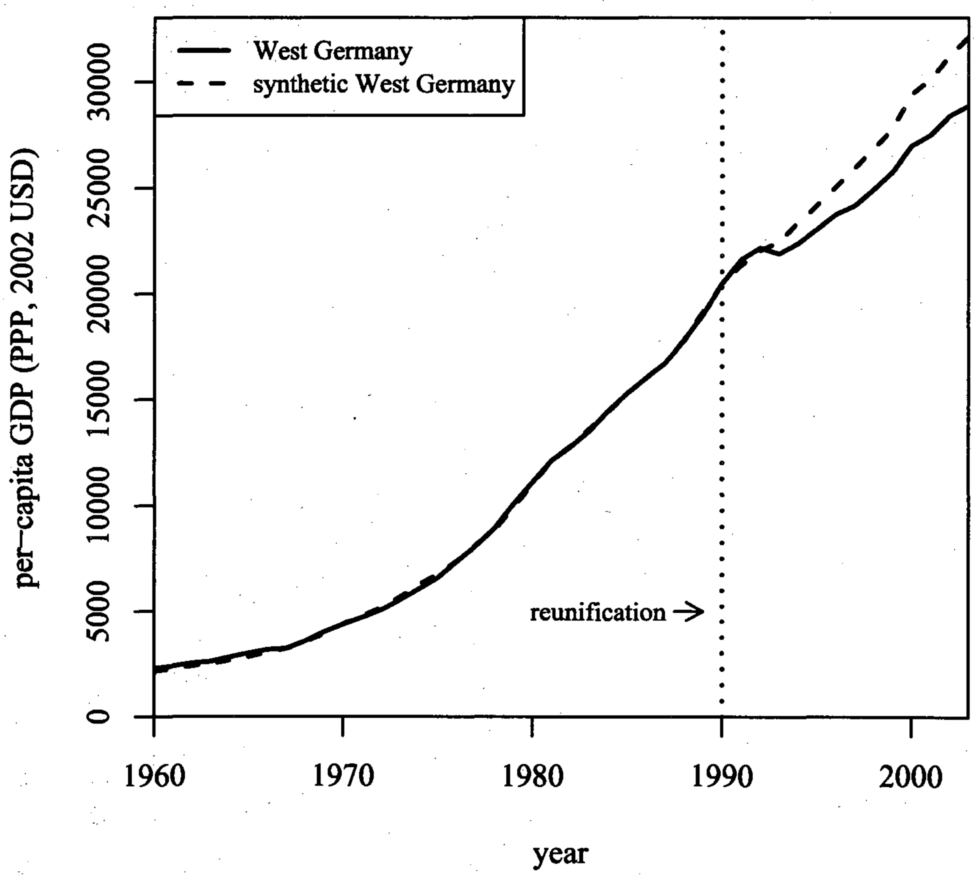
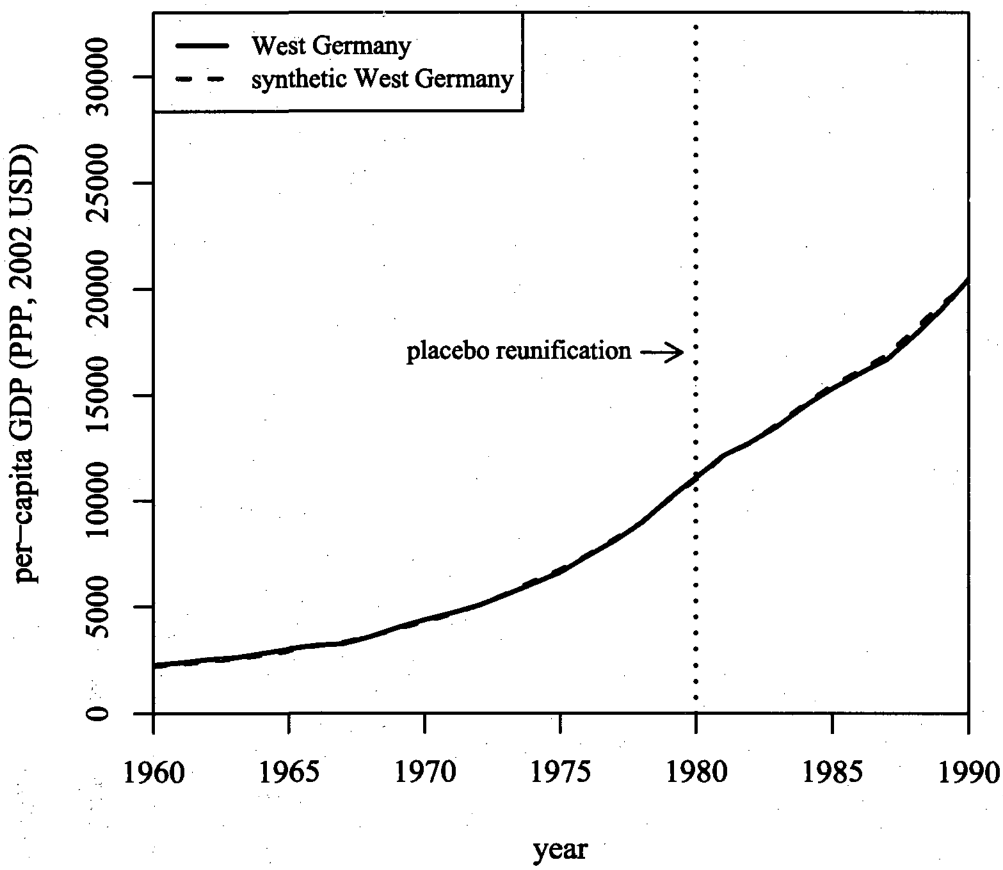


Figure 2.10: Placebo in Time. Trends in Per-Capita GDP: West Germany vs. Synthetic West Germany



Chapter 3

Synthetic Matching for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies

3.1 Introduction

Matching methods are nowadays often used in observational studies to preprocess the data prior to the estimation of binary treatment effects under the assumption of selection on observables (for reviews see Imbens & Wooldridge (2008); Rubin (2006); Imbens (2004); Rosenbaum (2002*b*)). All matching methods share a common goal: to improve the balance of pretreatment characteristics across the treatment and control group, thereby making the two sets of covariate distributions more similar. Improving balance has long been recognized as desirable because balance tends to reduce the model dependency of standard estimators such as regression and related approaches (Abadie & Imbens 2007, Ho, Imai, King & Stuart 2007, Abadie & Imbens 2006).

Despite their recent popularity, matching methods still suffer from a variety of problems and there is no consensus view on how matching is best conducted. In this study, we address three key problems that are shared by most commonly used matching methods. First, while researchers tend to agree that the goal of matching is to achieve the highest possible degree of covariate balance, most matching methods fail to directly focus on this task. In the current best practice, researchers "manually" iterate between propensity score modeling, matching, and balance checking until they attain a satisfactory matching solution (Iacus, King & Porro (2008) provide a notable exception). The hope is that an accurately estimated propensity score will stochastically balance the covariates, but this requires finding the correct model specification and often fairly large samples. As a result of this intricate search process, low balance levels prevail in many studies (Diamond & Sekhon 2006). Even worse, matching may counteract bias reduction for the subsequent treatment effect estimation when improving balance on some covariates decreases balance on some other covariates (see Iacus et al. (2008) and Ho et al. (2007) for similar critiques).

A second problem with commonly-used matching methods is that they ignore valuable information about features of the covariate distributions. For example, a well estimated propensity score may stochastically equate the covariate means in the treatment and the matched control group. But if the means still differ after the matching, then auxiliary information about the known sample moments was left "on the table" in the sense that the moment conditions (i.e. equal means) are not incorporated into the matching. One can improve the matching by choosing a weighting scheme that adjusts to the known means directly. This is analogous to a survey researcher that obtains a sample from some target population and also knows some features of this target population (e.g. population totals from the census). In this case the survey researcher can always improve her inferences (both in terms of bias and efficiency) about *unknown* features of the target population by adjusting the sample to the *known* features of the target population.

The third problem that we will address is that most commonly-used matching methods rely on unnecessarily rigid weighting schemes that constrain the units weights to be either zero or one – a unit is either matched or discarded.[1] But one can often generate better balanced samples by freeing the unit weights to vary smoothly across units. This greatly increases the space of potentially good matching solutions and maximises the use of available information. Units close to the nearest neighbors are not discarded, but simply down-weighted appropriately. Propensity score weighting (PSW) does not impose such rigid weight constraints. Perhaps not surprisingly this technique has excellent theoretical properties (if the true propensity score is correctly estimated) as it reaches the semi-parametric efficiency bound for the estimation of causal effects (Hirano, Imbens & Ridder 2003). One potential downside of PSW is that it may be sensitive to misspecification of the propensity score model which is commonly estimated with a logistic or probit regression.

In this paper we introduce synthetic matching as a technique to create balanced samples for the (subsequent) estimation of treatment effects. Synthetic matching addresses all three weaknesses of conventional matching methods because it directly incorporates balance into the weight function applied to the units, frees the weights to vary smoothly across units, and fully exploits auxiliary information about the known sample moments. In contrast to conventional matching methods, synthetic matching begins with a set of balance constraints that the researcher directly imposes on the unit weights. The balance constraints are flexibly specified with a potentially large set of moment conditions that imply that the moments of the reweighted groups match exactly. After the researcher has chosen her desired level of covariate balance, synthetic matching searches for the set of weights that satisfies the balance constraints, but remains as close as possible (in an entropy sense) to a set of base weights consisting of either uniform weighting or weights provided by an initially esti-

[1] In pair matching the weights can also be positive integers when matching with replacement. In the case of 1 to B matching ($B \in \mathbb{I}^+$) the weights are usually $1/B$ for all matched units; the same applies in situations where two matches are tied.

mated propensity score. This recalibration of the unit weights adjusts for systematic and random inequalities in representation between the two groups to reduce bias, but at the same time moves the weights as little as possible to retain efficiency. Synthetic matching thereby creates reweighted or so-called "synthetic" groups that are identical on a potentially large set of sample moment characteristics while avoiding the loss of valuable information.[2]

Synthetic matching makes it easy for the researcher to obtain a very high degree of covariate balance by imposing a potentially large set of balance constraints. The calibrated weights that result from synthetic matching are guaranteed to satisfy the balance constraints exactly, so that the reweighted groups are essentially identical with respect to the first, second, or higher moments of the covariate distributions (including higher order interactions). Synthetic matching always (at least weakly) improves upon the balance that can be obtained by conventional matching adjustments with respect to the specified balance constraints. Balance checking in the conventional sense is no longer necessary for the characteristics included in the weight function since balance is obtained by construction of the moment constraints. The higher degree of balance translates into less model dependency for the estimation of treatment effects.

The weights that result from synthetic matching can be used with an inverse probability weighting estimator just like conventional propensity score weights that may be obtained from estimating the propensity score with a logistic regression. The synthetic matching based weights are preferable because they incorporate the knowledge about the known sample moments and thus balance the moments exactly in finite samples. Synthetic matching weights are also applicable to any other outcome model that researchers want to use after the weighting adjustment. For example, estimation may be further refined by combining the synthetic matching weights with

[2]The attribute "synthetic" borrows from Abadie & Gardeazabal (2003) and the work presented in the previous chapter which presents synthetic control methods for longitudinal comparative case studies. In these papers a single synthetic control unit is created as a weighted average of several control units to match the pretreatment trajectory and characteristics of the treated units. This differs in many respects from synthetic matching as developed in this paper, which is designed to created balanced samples to conduct causal inference with a binary treatment.

a weighted least squares regression of the outcome on the covariates resulting in a double robust estimator, first suggested by Robins & Rotnitzky (1995).

We use Monte Carlo simulations to demonstrate the desirable properties of synthetic matching in a variety of commonly-used benchmark settings. The simulations show that in finite samples, synthetic matching improves in root mean squared error (MSE) upon PSW even when the propensity score model (estimated from a logistic or probit regression) is correct. It dramatically improves upon PSW in bias and MSE if the propensity score is (slightly) misspecified. Synthetic matching also outperforms (in a MSE sense) various widely used matching techniques such as Mahalanobis distance matching, Genetic matching, propensity score matching, or joint Mahalanobis distance and propensity score matching. We present an empirical application to the LaLonde (1986) dataset, where synthetic matching almost perfectly adjusts the means and variances of 52 covariate combinations (including the raw variables, squared terms, and first-order interactions). This constitutes a much higher level of balance than achieved by any of the numerous substantive and methodological papers devoted to this data-set (according to commonly-used balance metrics). The treatment effect estimate falls close to the experimental benchmark, and the estimate is slightly more efficient than in previous studies.

We provide companion software that implements the methods proposed in this paper. Our synthetic matching algorithm has globally quadratic convergence so that it finds the solution weights very fast even in high dimensional datasets.

While synthetic matching provides an reweighting scheme explicitly for the context of causal inference in observational studies with a binary treatment (where the goal is to equate the covariate distributions across the treatment and the control group), important links exist between the reweighting scheme employed in synthetic matching and various strands of literatures in econometrics and statistics. In particular, the survey literature contains several reweighting schemes which are used to adjust sampling weights so that sample totals match population totals known from auxiliary data (for a recent review see Särndal & Lundström (2006)). More broadly, reweighting schemes are also widely used in the literature on methods of moments estimation, empirical likelihood, and exponential tilting, where moment conditions (often derived from economic theory) are commonly used to efficiently estimate structural parameters (Hansen 1982, Qin & Lawless 1994, Kitamura & Stutzer 1997, Imbens 1997, Imbens, Spady & Johnson 1998, Hellerstein & Imbens 1999, Owen 2001, Schennach 2007).

3.2 Setup of the Problem

3.2.1 Causal Inference in Observational Studies with Binary Treatments

Following the conventional setup of the Rubin Causal model we consider a random sample of $n = n_1 + n_0$ units drawn from a population of size $N = N_1 + N_0$, where $n \leq N$ and N_1 and N_0 refer to the size of the target population of treated units and the source population of control units respectively. Each unit is exposed to a binary treatment $D_i \in \{1, 0\}$; $D_i = 1$ if unit i received the active treatment and $D_i = 0$ if unit i received the control treatment. In the sample we have n_1 treated units and n_0 control units. Each unit has a vector of exogenous pretreatment characteristics $X_i = [X_{i,1}, X_{i,2}, ..., X_{i,k}]$ with $k \geq 1$ and $i \in \{1, 2, ..., n\}$. Let $f_{X|D=1}$ and $f_{X|D=0}$ denote the densities of these covariates in the treatment and control population respectively. Finally, let $Y_i(W_i)$ denote the pair of potential outcomes that individual i attains if it is exposed to the active treatment or the control treatment.[3] Observed outcomes for each individual are realized as $Y_i = Y_i(1)\, D_i + (1 - D_i)Y_i(0)$ so that for each unit we never observe both potential outcomes simultaneously but the triple (D_i, Y_i, X_i).

The unobserved treatment effect for each unit is defined as $\tau_i = Y_i(1) - Y_i(0)$. The quantities of interest are defined as functions of τ_i for different subsets of units. Most common are the sample (SATE) and population (PATE) average treatment effects given by SATE $= n^{-1} \sum_i^n \tau_i$ and PATE $= N^{-1} \sum_i^N \tau_i$ and the sample (SATT)

[3]Notice that this implies the stable unit treatment value assumption (SUTVA) which rules out interference of outcomes among units and non-identical versions of the treatment (Rubin 1978).

and population (PATT) average treatment effect on the treated given by SATT $= n_1^{-1}\sum_{\{i|D=1\}}^{n}\tau_i$ and PATT $= N_1^{-1}\sum_{\{i|D=1\}}^{N}\tau_i$. Notice that $\mathbb{E}[\text{SATE}] = \text{PATE} = \mathbb{E}[Y(1)-Y(0)]$ and similarly $\mathbb{E}[\text{SATT}] = \text{PATT} = \mathbb{E}[Y(1)-Y(0)|D=1]$ since we consider random samples. In the following we focus on the PATT as our central quantity of interest. The synthetic matching methods introduced below are equally applicable to estimate the other commonly-used quantities of interest analogously.[4]

Our goal is to estimate $\tau = \mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=1]$. The first term in the expression, $\mathbb{E}[Y(1)|D=1]$, is easily estimable from the data. The second expression, $\mathbb{E}[Y(0)|D=1]$, is counterfactual and thus unobserved in the target population. The only information available about $Y(0)$ is in the source population not exposed to the treatment. In experimental studies, where treatment assignment is forced to be independent of the potential outcomes, $Y(1), Y(0) \perp D$, we can simply use $\mathbb{E}[Y(0)|D=0]$ as our estimate of $\mathbb{E}[Y(0)|D=1]$. In observational studies, however, selection into treatment usually renders the latter two quantities unequal. The conventional solution to this problem is to assume ignorable treatment assignment and overlap (Rosenbaum & Rubin 1983), which implies that $Y(0) \perp D|X$ and that $\Pr(D=1|X=x) < 1$ for all x in the support of $f_{X|D=1}$. Therefore, conditional on all confounding covariates X the potential outcomes are stochastically independent of D and the PATT is identified as:

$$\tau = \mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=1] = \mathbb{E}[Y|D=1] - \int \mathbb{E}[Y|X=x, D=0] f_{X|D=1}(x)dx$$

where the integral is taken over the support of X in the source population. Notice that the last term in this expression is equal to the covariate adjusted mean, i.e. the estimated mean of Y in the source population (the control group) if its covariates were distributed as in the target population (the treatment group).

To intuitively see why covariate balance is key for the estimation of the PATT, notice that the potential outcomes for the treated units can be written as $Y_i(D_i) = l(X_i)$ where $l()$ is an unknown function. For simplicity suppose that the treatment effect is estimated by the difference in means. The true treatment effect can then be

[4]E.g., the treatment group can be reweighted to match the control group.

decomposed into the estimated treatment effect and the average estimation error:

$$\text{PATT} = \widehat{\text{PATT}} + N_1^{-1} \sum_{\{i|D=1\}} \left(l_0(X_{\{i|D=0\}}) - l_0(X_{\{i|D=1\}})\right)$$

where $l_0(X_{\{i|D=0\}}) - l_0(X_{\{i|D=1\}}) = \hat{Y}_i(0) - Y_i(0)$ is the unit level treatment error (see Iacus et al. (2008)). The estimation error has two components: i) the unknown function $l()$, which determines the importance of the variables, and ii) the imbalance, which is defined as the difference between the empirical covariate distribution of the treatment $f_{X|D=1}$ and the control group $f_{X|D=0}$. The goal of matching is to reduce the imbalance to decrease the estimation error as far as possible.[5]

A variety of estimators have been proposed for $\mathbb{E}[Y|X = x, D = 0]$ (see Imbens (2004) and Rubin (2006) for extensive reviews). For example, if X has only a few dimensions, $\mathbb{E}[Y|X = x, D = 0]$ is often estimated using local linear regression, pair matching, or nearest neighbor matching. If X has many dimensions, non-parametric regression is often very difficult. As a solution, the problem can be reduced to a single dimension. As was first shown by Rosenbaum & Rubin (1983) and Rosenbaum & Rubin (1984) the counterfactual mean can also be identified as:

$$\mathbb{E}[Y(0)|D = 1] = \int \mathbb{E}[Y|p(X) = \rho, D = 0] f_{p|D=1}(\rho) d\rho$$

where $f_{p|D=1}$ is the distribution of the propensity score $p(x) = \Pr(D = 1|X = x)$ in the target population. This follows from their result that $Y(0) \perp D|X$ implies $Y(0) \perp$

[5]One important aspect is that there exists no agreement on how to assess covariate balance in practice. Theoretically, we would like the two empirical distributions to be equal so that the density in the matched control group $f^*_{X|D=0}$ mirrors the density in the treatment group $f_{X|D=1}$. Balance checking would then involve a comparisons of the (joint) empirical distributions of all covariates X. But this is infeasible when X is high dimensional and in practice lower dimensional measures of balance are used (but see Iacus et al. (2008) who propose a multidimenional balance metric). Opinions differ on what metric is most appropriate. The most commonly-used metric is the standardized difference in means (Rosenbaum & Rubin 1983) and t-tests for differences in means. Diamond & Sekhon (2006) argue that paired t-test and bootstrapped Kolmogorov-Smirnov (KS) tests should be used instead and that commonly-used p-value cutoffs such as .1 or .05 are too lenient to obtain reliable causal inferences. Rubin (2001) also considers variance ratios and tests for residuals that are orthogonalized to the propensity score. Imai, King & Stuart (2008) criticize the use of t-tests and stopping rules and argue that all balance measures should be maximised without limit. They advocate QQ plot summary statistics as better alternatives than t-tests or KS tests. Sekhon (2006) comes to the opposite conclusion. No scholarly standard is in sight. A key advantage of synthetic matching is that it obviates balance checking altogether at least for the moments included in the balance constraints.

$D|p(x)$. A variety of propensity score (matching) estimators have been proposed to estimate the counterfactual mean outcome of the treated in the absence of the treatment.

3.2.2 Propensity Score Weighting (PSW)

The estimator of particular interest in this paper is based on weighting on the propensity score as suggested in the seminal papers by Hirano & Imbens (2001) and Hirano et al. (2003) (also see Imbens (2004)). Here the counterfactual mean is estimated as

$$\widehat{\mathbb{E}[Y(0)|D=1]} = \frac{\sum_{\{i|D=0\}} Y_i\, d_i}{\sum_{\{i|D=0\}} d_i} \tag{3.1}$$

where every control unit receives a weight given by $d_i = \frac{\hat{p}(x_i)}{1-\hat{p}(x_i)}$. The structure of this estimator is similar to that of the classic Horvitz-Thompson estimator used in the survey literature where units are weighted by the inverse of the inclusion probabilities that result from the sampling design (Horvitz & Thompson 1952). In PSW units are weighted by the inverse of the (estimated) assignment probabilities given by the propensity score. This similarity between survey sampling weights and propensity score weights provides the entry point for the reweighting methods introduced below.

Despite its fairly recent development, PSW is already widely used in a variety of fields.[6] This may not be surprising given that PSW has some very attractive features compared to other adjustment techniques such as pair matching or propensity score matching. Hirano et al. (2003) show that PSW achieves the semiparametric efficiency bound for the estimation of average causal effects (Robins, Rotnitzky & Zhao 1995, Hahn 1998). This result requires sufficiently large samples and a propensity score that is sufficiently flexibly estimated to closely approximate the true propensity score. Intuitively, if the propensity score is well estimated, the control observations are reweighted so that they form a balanced sample with the treated

[6]For examples in economics see Eliason & Storrie (2006) and van de Walle & Mu (2007); in eduction see Frank, Sykes, Anagnostopoulos, Cannata, Chard, Krause & McCrory (2008); in public health see Elliott, Beckett, Chong, Hambarsoomians & Hays (2008) and Indicators & Analyses (2007); in psychology see McNiel & Binder (2007); in criminology Ridgeway (2006).

observations.[7] If applied correctly, PSW is more efficient in finite samples than other non-parametric adjustment techniques such as matching on the covariates or matching on the propensity score.[8]

3.2.3 Achieving Balance with Matching Methods

While in theory PSW provides a very attractive option for researchers, it suffers from the same practical drawbacks that plague all propensity score methods: the true propensity score is usually unknown and needs to be estimated in the first step. Propensity score methods allow us to estimate treatment effects without explicitly modeling the outcome conditional on covariates and treatment status, but they still require an equally high-dimensional model of the assignment probability conditional on the covariates. Several studies have demonstrated that misspecified propensity scores can lead to substantial bias for the estimation of treatment effects (Drake 1993, Smith & Todd 2001, Smith & Todd 2005*a*, Smith & Todd 2005*b*, Diamond & Sekhon 2006).

In the current best practice, researcher try to avoid propensity score misspecification by "manually" iterating between matching, propensity score modeling, and balance checking until a satisfactory matching solution is reached. This follows the advice of Rosenbaum & Rubin (1983) who describe the propensity score as primarily a balancing score: the resulting balance provides the appropriate yardstick to assess the accuracy of a propensity score model. Some researchers have criticized this cyclical process as the "propensity score tautology" (Imai et al. 2008).

[7] Formally propensity score reweighting exploits the following equalities: $\mathbb{E}[\frac{DY}{p(x)}] = \mathbb{E}[\frac{DY(1)}{p(x)}] = \mathbb{E}[\mathbb{E}[\frac{DY(1)}{p(x)}|X]] = \mathbb{E}[\frac{p(x)Y(1)}{p(x)}] = \mathbb{E}[Y(1)]$ which uses the ignorability assumption in the second to last equality (Hirano & Imbens 2001, Hirano et al. 2003).

[8] Notice that weighting by the estimated rather than the true propensity score in general leads to better balance and thus more efficient estimates (Robins et al. 1995, Rubin & Thomas 1996, Hahn 1998, Hirano et al. 2003). The intuitive reason, as explained in Rosenbaum (1987, pg. 391), is that weighting by the true propensity scores "compensates only for the systematic differences" between groups while weighting by an estimated propensity scores "[corrects] for both systematic and chance imbalances." See also Hirano et al. (2003) who derive this result for the weighting estimator using an empirical likelihood interpretation.

More importantly, when estimating propensity scores in practice it is often difficult to jointly balance all covariates. The iterative process of tweaking the propensity score model and balance checking can be tedious and frequently results in low balance levels. Even worse, as Diamond & Sekhon (2006, pg. 8) observe, a "significant shortcoming of common matching methods such as Mahalanobis distance and propensity score matching is that they may (and in practice, frequently do) make balance worse across measured potential confounders." Unless the distributions of the covariates are ellipsoidally symmetric (Rubin & Thomas 1992) or are mixtures of proportional ellipsoidally symmetric distributions (Rubin & Stuart 2006), there is no guarantee that the matching techniques will be equally percent bias reducing (EPBR) (Rubin 1976*a*, Rubin 1976*b*). Therefore the bias of some linear functions of X may be increased while all univariate covariate means are closer after matching.[9]

Also notice that even with a good propensity score model, imbalances often remain because stochastic balancing occurs only asymptotically. Chance imbalances may remain in any finite sample (see Hirano & Imbens (2001) for an example in the context of propensity score weighting) and in these cases one can still improve the balance by enforcing balance constraints on the specified moments.

One way to address the problem of low balance is to resort to better estimation techniques for the propensity score model. In practice, researchers almost always rely on simple parametric models such as logistic regression (as originally proposed by Rosenbaum & Rubin (1983)), but better models are available. For example, Hirano et al. (2003) derive their celebrated PSW result for a case where the propensity score is estimated using a nonparametric series estimator that approximates the true propensity score by a power series in all variables. Asymptotically, this series will converge to the true propensity score function if the powers increase with the sample size, but no results exist about the finite sample properties of this estimator. By the authors' own admission this approach is computationally not very attractive. Lately other estimation methods such as boosted regression (McCaffrey, Ridgeway &

[9]Ellipsoidal symmetry fails when for example the covariate set includes binary, categorical, and or skewed continuous variables.

Morral 2004) or kernel regression (Frölich 2007) have been suggested.

We argue that a more promising proposal is to focus on balance directly. An important first step in this direction is the idea of Genetic matching as developed by Diamond & Sekhon (2006). Genetic matching finds nearest neighbors based on a generalized Mahalanobis distance with variable covariate weights. The covariate weights are chosen by a genetic algorithm in order to find a set of matched units that maximise covariate balance measured by the minimum p-value across a set of balance tests. By incorporating balance criteria into the weight function applied to the covariates, Genetic matching can improve balance compared to other matching techniques (see the simulations in Diamond & Sekhon (2006)).

Synthetic matching differs from genetic matching in several important aspects. While Genetic matching searches for covariate weights that lead to a matching that balances the covariates, synthetic matching searches for *unit* weights that balance the covariates directly. It begins with the specification of balance constraints and then searches for a set of unit weights that satisfy the specified level of balance while moving the weights as little as possible. This obviates the need for balance checking altogether, at least with respect to the moments included in the balance constraints. In other words, if a solution exists, synthetic matching achieves exact balance on the sample moments, so the reweighted control group and the treatment group have identical means, variances and specified higher moments. Moreover, by freeing the weights to vary smoothly across units synthetic matching also gains efficiency as it dispenses with the rigid weight constraints that require that a unit is either matched (weight of one) or discarded (weight of zero). Finally, since the optimization problem involved in synthetic matching is well behaved and globally convex, the solution is usually obtained within seconds even in large datasets. In Genetic matching the optimization problem is difficult and irregular and a genetic algorithm is used to maximise balance.

3.3 Estimating Treatment Effects by Synthetic Matching

In the following we motivate synthetic matching by introducing an estimator for the PATT $\tau = \mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 1]$ where we only need to estimate the counterfactual mean $\mathbb{E}[Y(0)|D = 1]$. Notice that we focus on the counterfactual mean here purely for convenience. Other commonly-used quantities of interest can be similarly estimated since the calibrated set of w_i that result from synthetic matching identifies the entire counterfactual distribution of $Y(0)$ in the target population. Moreover, it is important to recognize that synthetic matching is primarily a method of covariate adjustment to create balanced samples. The resulting weights can be combined with any outcome model of interest that the researcher may want to use to model the relationship between the outcomes and the covariates in the matched dataset to address remaining imbalances between the two groups. The inverse probability weighting estimator of the treatment effect we present here refers to the simplest possible choice of such an outcome model, namely the weighted difference in mean outcomes between the treated units and the reweighted control units.

3.3.1 Synthetic Matching

Formally, the synthetic matching estimator for the counterfactual mean is given by

$$\mathbb{E}[Y\widehat{(0)|D} = 1] = \frac{\sum_{\{i|D=0\}} Y_i w_i}{\sum_{\{i|D=0\}} w_i} \tag{3.2}$$

where a scalar weight w_i is chosen for each control unit by the following reweighting scheme

$$\min_{w_i} H(w) = \sum_{\{i|D=0\}} h(w_i) \tag{3.3}$$

subject to balance and normalizing constraints:

$$\sum_{\{i|D=0\}} w_i\, c_{ri}(X_i) = m_r \quad \text{with} \quad r \in 1, ..., R \quad \text{and} \tag{3.4}$$

$$\sum_{\{i|D=0\}} w_i = 1 \quad \text{and} \tag{3.5}$$

$$w_i \geq 0 \quad \text{for all} \quad i \quad \text{such that} \quad D = 0. \tag{3.6}$$

The estimator consists of three features. First, the loss function $h(\cdot)$ is a distance metric chosen from the general class of empirical minimum discrepancy estimators defined by the Cressie-Read (CR) divergence (Read & Cressie 1988). We prefer to use the directed Kullback entropy divergence (Kullback 1959) defined by $h(w_i) = w_i log(w_i/q_i)$ with estimated weight w_i and base weight q_i.[10] The loss function measures the distance between the distribution of estimated control weights defined by the vector $W = [w_i, ..., w_{n_0}]'$ and the distribution of the base weights specified by the vector $Q = [q_i, ..., q_{n_0}]'$. Notice that the loss function decreases the closer W is to Q; the loss equals zero if $W = Q$. We use the set of uniform weights with $q_i = 1/n_0$ as our base weights. However, other base weights may be utilized to include prior information such as weights from an initially estimated propensity score that are then further adjusted by the reweighting scheme to accommodate the specified balance constraints (see below).

Second, the R constraints defined in 3.4 we refer to as our balance constraints. They are imposed by the researcher to equalize the moments of the covariate dis-

[10] The Cressie-Read divergence family may be described by $h(w) \equiv CR(\gamma) = \frac{w^{\gamma+1}-1}{\gamma(\gamma+1)}$ where γ indexes the family and limits are defined by continuity so that $\lim_{\gamma\to 0} CR(\gamma) = \lim_{\gamma\to 0} \frac{w^{\gamma+1}-1}{\gamma} = \lim_{\gamma\to 0} wlog(w)$ and $\lim_{\gamma\to -1} CR(\gamma) = \lim_{\gamma\to -1} \frac{w^{\gamma+1}-1}{\gamma} = \lim_{\gamma\to -1} - log(w_i)$ where the last equalities follow from l'Hospital respectively. Notice that $h(w) = wlog(w)$ represents the Shannon entropy metric which is (up to a constant) equivalent to the Kullback entropy divergence when uniform weights q_i are used for the null distribution. The estimator is valid for other choices of γ. Another choice with good properties is $\gamma = -1$ which results in an empirical likelihood (EL) estimator. We prefer the entropy estimator because it is more robust under misspecification (Imbens et al. 1998, Schennach 2007, Baggerly 2003)). Another desirable feature of the entropy estimator is that it constrains the weights to be zero or positive, while the EL allows for negative weights. For general reviews on the choices see Qin & Lawless (1994), Owen (2001, ch.3), Imbens (1997), Newey & Smith (2004), or Schennach (2007).

tribution between the treatment and the reweighted control group. A typical constraint is formulated with m_r containing the r-th order moment of a given variable X_j from the target population (ie. the treatment group) while the moment functions are specified for the source population (ie. the control group) as $c_{ri}(X_{ij}) = X_{ij}^r$ or $c_{ri}(X_{ij}) = (X_{ij} - \mu_j)^r$ with mean μ_j.

Third, the constraints in 3.5 we refer to as our normalization constraints. The first condition implies that the weights sum to the normalization constant of 1. This choice is arbitrary and other constants can be used.[11] The second condition implies a non-negativity constraint because the distance metric is not defined for negative weight values. Below we see that this constraint is not-binding and can be safely ignored.

The rationale behind this reweighing scheme is to reweight the control group by finding a set of weights that is as close as possible (in an entropy sense) to the set of uniform base weights but nonetheless consistent with the balance constraints that imply that the moments in the reweighted control group match the moments in the treatment group. In other words, the researcher first specifies a desired level of balance using a potentially large set of balance constraints. Then the estimator searches for a set of weights that guarantees this pre-specified level of balance but retains the maximum information in the control group by keeping the weights as close as possible to uniform weighting. The key advantage is that balance checking is no longer necessary since the balance constraints are directly built into the weight function applied to the units and exact balance on the specified moments will be obtained in finite samples.

In the case of a large randomized experiments where the distributions are (asymptotically) balanced before the reweighing, the specified balance constraints in 3.4 are non-binding (assuming no chance imbalances) and the counterfactual mean is estimated as a weighted average of the control unit outcomes with every control unit

[11] The conventional propensity score weights $p/(1-p)$ for the average treatment effect on the treated imply that the sum of the control weights is equal to the number of treated units which is identical to setting the normalization constraint to n_0 instead of 1 in 3.5. The companion software allows the researcher to set the normalization constant.

weighted equally. The higher the level of imbalance in the covariates distributions, the further the weights have to be adjusted to meet the balance constraints. The number of moment conditions may vary depending on the dimensionality of the covariate space, the shapes of the covariate densities in the two groups, the sample sizes, and the desired balance level. At a minimum the researcher would want to adjust at least the first moments of the marginal distributions of all confounders in X, but variances can be similarly adjusted. In most empirical cases we would expect the bulk of the confounding to depend on first and second moments. If, however, the researcher is concerned about dependencies in higher moments these can be similarly adjusted to by including higher moments in the condition vector. Interactions can be similarly included. The number of moment constraints can be increased at a constant rate with a growing sample size.

Notice that this reweighing scheme is analogous to reweighting adjustments that are sometimes used in the survey literature to correct sampling weights for bias due to non-response, frame undercoverage, response biases or to integrate auxiliary information to improve precision of estimates. The idea is that by introducing auxiliary information about known characteristics of the target population (for example population totals known from the census) one can improve estimates about unknown characteristics of the target population by adjusting the sampling design weights so that the sample moments match (at least) the known population moments. These adjustments include a wide variety of methods such as post-stratification, raking, and calibration estimators (see for example Deming & Stephan (1940), Oh & Scheuren (1978) or Särndal & Lundström (2006) for a recent review). Zaslavsky (1998) proposes a similar log-linear reweighting scheme with an entropy divergence to adjust for undercount in census data. Ireland & Kullback (1968) develop a minimum discrimination estimator that fits the cell probabilities of a (multidimensional) contingency table based on fixed marginal probabilities by minimizing the directed entropy divergence (starting from equal weights). They show that minimizing the entropy provides a BAN estimator that is consistent as well as asymptotically normal and efficient.

In contrast to most applications of reweighting in a survey context, where the vector of auxiliary information is commonly limited to a few totals, in the case of synthetic matching the data from the treatment group allows us to create a very large set of moment conditions. This forces the density of X in the reweighted control group to look very close to that in the treatment group. Moreover, by including balance constraints for all confounders the researcher can rule out the possibility that balance is decreased on some observed confounders.

3.3.2 Implementation

To fit the weights implied by the synthetic matching estimator, we need to minimize the loss function $h(w_i) = \sum_{\{i|D=0\}} w_i \, log(w_i/q_i)$ subject to the balance and normalization constraints given in equations 3.4 and 3.5. Using the Lagrange multiplier we obtain the primal optimization problem:

$$\min_{W,\lambda_0,Z} L^p = \sum_{\{i|D=0\}} w_i \, log(w_i/q_i) + \sum_{r=1}^{M} \lambda_r \Big(\sum_{\{i|D=0\}} w_i c_{ri}(X_i) - m_r \Big) + (\lambda_0 - 1)\Big(\sum_{\{i|D=0\}} w_i - 1 \Big) \tag{3.7}$$

where $Z = \{\lambda_1, ..., \lambda_R\}'$ is a vector or Lagrange multipliers for the balance constraints and $\lambda_0 - 1$ the Lagrange multiplier for the normalization constraints. This system of equations is computationally inconvenient given its dimensionality of $n_0 + R + 1$. However, we can exploit several structural features that make this problem very susceptible to solution. First, the loss function is strictly convex since $\frac{\partial^2 h}{\partial W^2} > 0$ for $w_i \geq 0$, so that every local solution W^* is a global solution and any global solution is unique if the constraints are consistent. Second, as was first recognised by Erlander (1977), Wolfe duality holds and we can substitute out the constraints.[12] The first order condition of $\frac{\partial L^p}{\partial w_i} = 0$ yields that the solution for each weight is attained by:

$$w_i^* = \frac{q_i exp(-\sum_{r=1}^{R} \lambda_r c_{ri}(X_i))}{\sum_{\{i|D=0\}} q_i exp(-\sum_{r=1}^{R} \lambda_r c_{ri}(X_i))} \tag{3.8}$$

[12] Also see the excellent reviews on entropy optimization in Eriksson (1980), Fletcher (1987), Kapur & Kevsavan (1992), Milman, Jiang & Jelliffe (2001), Mattos & Veiga (2004))

The expression makes clear that the weights are estimated as a log-linear function of the covariates specified in the moment conditions.[13] Plugging this expression back into L^p eliminates the constraints and leads to the unrestricted dual problem given by:

$$\max_{Z} L^d = log\left(\sum_{\{i|D=0\}} q_i exp(-\sum_{r=1}^{R}\lambda_r c_{ri}(X_i))\right) - \sum_{r=1}^{R}\lambda_r m_r \tag{3.9}$$

The solution to the dual problem Z^* solves the primal problem and the weights W^* can be recovered via 3.8. This dual problem is much more tractable because it is unconstrained and dimensionality is reduced to only R. Moreover, a unique solution exists as L^d is strictly convex.

We use a Levenberg-Marquardt scheme to find Z^* for this dual problem. We rewrite the constraints in matrix form by defining the $(R \times n_0)$ constraint matrix $C = [c_{i1}(X_i), ..., c_{ir}(X_i)]'$ and the moment vector $M = [m_1, ..., m_R]'$, which allows the balance constraints to be written as $CW = M$. Notice that C' has to be of full column rank otherwise the system has no feasible solution because the constraints are not linearly independent. The problem can then be written as

$$\max_{Z} L^d = \log(Q' exp(-C'Z)) + M'Z \tag{3.10}$$

with solution $W^* = \frac{Q \cdot exp(-C'z))}{Q' exp(-C'Z)}$. The gradient and hessian are $\frac{\partial L^d}{\partial Z} = M - CW$ and $\frac{\partial^2 L^d}{\partial Z^2} = C[W'I - WW']C'$ where I is the n_0 dimensional identity matrix. We exploit this second order information by iterating

$$Z^{new} = Z^{old} - l\,\nabla_Z^2 L^{d^{-1}} \nabla_Z L^d \tag{3.11}$$

where l is a scalar that denotes the step length. In each iteration we either take the full Newton step $l = 1$ if that lowers the objective function and if not l is chosen by backtracking in the Newton direction to the optimal step length using line search that combines a golden section search and successive quadratic approximation. We use $Z^0 = (CC')^{-1}M$ as our starting guess. This iterative algorithm is globally convergent if the problem is feasible, and the solution is usually obtained in minimal computing

[13]Evidently, the inequality bounds $w_i \geq 0$ are inactive and can be safely ignored.

time after a few iterations. Companion software that implements the estimator is available from the author.

3.3.3 Estimated Propensity Score as Base Weight

While the previous discussion assumed the case of minimizing the distance from uniform weights $q_i = 1/n_0$ the synthetic matching adjustment may be used with other set of base weights. In the survey world, the base weight usually comes from the sampling design and the goal is to adjust the sample by reweighting while moving the design weights as little as possible (Oh & Scheuren 1978, Zaslavsky 1998, Särndal & Lundström 2006). In our context a natural base weight may be a weight constructed from a propensity score that is initially estimated with a logistic regression. These base weights provide a first step towards balancing the covariates. Synthetic matching then "overhauls" the weights to fix remaining (chance) imbalances in the specified moments but keeps the weights as close as possible (in a relative entropy sense) to the original propensity score weights. If the propensity score is correctly specified and the sample size grows large, the first step estimation will be sufficient to stochastically balance the moments between the groups. In this case the balance constraints are non-binding and the synthetic matching adjustments will not alter the initially estimated propensity score weights.

3.3.4 Double Robust Estimation

Synthetic matching weights are easily combined with regression models that directly address the correlation between the covariates and the outcome by using weighted least squares. This combination leads to an estimator that is "double robust" in the sense that it remains consistent if either the weighting model for the treatment or the outcome model that relates the outcome to the covariates is correctly specified. The outcome model can also increase precision if the (additional) variables in the outcome model account for residual variation in the outcome of interest (Robins et al. 1995, Robins & Rotnitzky 1995, Robins & Ritov 1997, Hirano &

Imbens 2001).

3.4 Monte Carlo Simulations

To demonstrate the desirable properties of synthetic matching we present Monte Carlo experiments that rely on commonly-used benchmark settings.[14]

3.4.1 Competing Estimators

For each experiment, we compare the following commonly-used matching estimators:

- **Raw**: Refers to the unadjusted difference in means between the treated and the control group: $\beta_{Raw} = n_1^{-1} \sum_{\{i|D=1\}} Y_i - n_0^{-1} \sum_{\{i|D=0\}} Y_i$
- **Propensity Score Matching (PS)**: Refers to propensity score matching where each treated unit is matched with replacement to B nearest control units on the the propensity score metric. We estimate the propensity score using probit regression and match on the linear predictor $\hat{\mu} = X\hat{\beta}$ to avoid com-

[14] There is a growing literature that uses simulation to assess the properties of matching estimators (partially reviewed in Imbens (2004)). Frölich (2004) presents an extensive simulation study to consider various matching estimators under a variety of settings. While he considers a wide variety of sample designs, his study is limited to a single covariate and true propensity scores. Zhao (2004) investigates the finite sample properties of pair matching and propensity score matching and finds no clear winner among these techniques. While including different sample sizes, his study does not vary the controls to treated ratio and is also limited to true propensity scores. Brookhart, Schneeweiss, Rothman, Glynn, Avorn & Sturmer (2006) simulate the effect of including or excluding irrelevant variables in propensity score matching that are either correlated with the outcome or the treatment. Abadie & Imbens (2007) present a matching simulation using data from the PSID data and find that their bias corrected matching estimator outperforms linear regression adjustment. Diamond & Sekhon (2006) provide two monte carlo experiments one with multivariate normal data and three covariates and a second one using data from the Lalonde dataset. They find that their genetic matching outperforms other matching techniques. Further simulations using multivariate normal data are presented in Gu & Rosenbaum (1993) and several of the papers collected in Rubin (2006). Drake (1993) presents a simulation study to investigate sensitivity to the misspecified propensity scores in the case of two normally distributed covariates. He finds that misspecification often results in substantial bias.

pression of the propensity scores between zero and one. We vary the degree of misspecification of the propensity score model (details below).

- **Mahalanobis Matching (MD)**: Refers to Mahalanobis distance matching, where each treated unit is matched to B nearest neighbors according to the distance metric $d(X_i, X_j) = \{(X_i - X_j)'(S^{-1/2})'VS^{-1/2}(X_i - X_j)\}^{1/2}$ where V is a $(k \times k)$ positive definite weight matrix with zero in all elements except the main diagonal and $S^{1/2}$ is the Cholesky decomposition of S, the variance-covariance matrix of X. The k parameters in the diagonal of V are set equal to 1.

- **Genetic matching (GM)**: Refers to Genetic matching. The estimator is similar to MD Matching, except that the weights in the diagonal of W are chosen by an evolutionary algorithm such that balance across treatment and control groups is maximized (Diamond & Sekhon 2006). We use the default setting in the `GenMatch()` function in `R` (R Development Core Team 2009), where balance is measured by the lowest p-value across covariate-by-covariate paired t-tests for differences in means and bootstrapped Kolmogorov-Smirnov tests for the equality of distributions (see Sekhon (2007) for details).

- **PS + MD Matching (PSMD)**: Refers to a combination of PS and MD matching where each treated unit is matched to B nearest neighbors using a MD defined by the linear predictor of the estimated propensity score, $\hat{\mu}$ and the covariates X once they have been othorgonalized to the linear predicton as in Rubin (2001). Notice that for PS, MD, and PSMD matching we use the `Match()` function in `R` (Sekhon 2007).

- **PS Weighting (PSW)**: Refers to weighting on the estimated propensity score as described above.

- **Synthetic Matching (SM)**: Refers to the synthetic matching estimator as outlined above. Unless otherwise noted we adjust only first moments of the covariates to bias the results against the SM estimator.

All matching is conducted with replacement and $B = 1$. For all matching estimators, the counterfactual mean is computed as the (weighted) average of the matched control units.

3.4.2 Monte Carlo Experiment I: EPBR Data

Design

The first experiment is based on a setting that meets the conditions that are necessary for matching estimators to achieve the EPBR property (as in Rubin and Thomas (1992)). The design is inspired by the first experiment in Diamond & Sekhon (2006), but extends their analysis by adding two additional scenarios. We use 50 treated observations and 100 control observations, with three baseline covariates X that are multivariate normal with zero covariances. For the treated observations the means of the X variables are .20 and for the control observations the means are equal to 0. We generate outcomes with a linear mapping $Y = X\beta + \epsilon$ where $\epsilon \sim N(0, .5)$ and $\beta = (1, 1, 1)'$. The true treatment effect is zero for all units. We consider three variants of this designs:

- Design A: the variances of X are unity in both groups; the easiest, but presumably most unrealistic case.
- Design B: the variances of X are .5 in the control group and and 1.5 in the treatment group. This provides a somewhat more difficult but realistic case; in many empirical cases variances may differ between the two groups.
- Design C: the variances are equal to unity, but we include the three squared terms of the variables in X in the design matrix for the estimators. The squared terms are omitted from the outcome equation (ie. the mapping from X to Y) so that this scenario mirrors a case where a researcher adjusts for three irrelevant covariates (as in Brookhart et al. (2006)). We include this case because adjusting for squared terms is often recommended in practice.

For the propensity score methods we use the correctly estimated propensity score (from a logistic regression that is linear in X), instead of the true score because the former is known to be more efficient. We run 1000 simulations and report the mean estimate which indicates the bias (multiplied by 100) and the root mean squared error (MSE).

Results for Design A: Equal Variances

The results are shown in the upper panel of table 3.1. We find that the synthetic matching estimator is unbiased and highly efficient; it outperforms all other propensity score methods and multivariate matching methods in terms of MSE. In particular, the synthetic matching estimator has an MSE that is more than four times lower than that of the conventional propensity score weighting estimator where a probit regression is used to estimate the score. This is expected because the synthetic matching estimator fully incorporates the information about the known sample moments. The MSE of synthetic matching is 13 times lower compared to propensity score matching, about 8 times lower than Mahalanobis matching and about 11 times lower than the joint propensity score Mahalanobis distance matching. Consistent with Diamond & Sekhon (2006) Genetic matching dominates the other matching techniques in terms of MSE, but its MSE is still more than 3 times larger than that of synthetic matching. The fact that the matching estimators (except Genetic matching) are generally less efficient is consistent with the results from Abadie & Imbens (2006). We also find that the multivariate matching methods are all biased. This is consistent with Abadie & Imbens (2006) who show that the bias of matching estimators is of order $O(N^{-1/k})$ where k is the number of continuous covariates.

Results for Design B: Unequal Variances

The results are shown in the middle panel of table 3.1. Synthetic matching dominates the other estimators and the differences in MSE are now amplified due to the higher variances in the treatment group.

Results for Design C: Irrelevant Covariates

The results are shown in the lower panel of table 3.1. The inclusion of the irrelevant variables does not adversely affect synthetic matching, but the other methods now exhibit lower MSE compared to design A. In summary, the first monte carlo experiment shows that in a finite sample where the conditions necessary for EPBR hold, synthetic matching clearly outperforms all other non-parametric adjustment techniques.

3.4.3 Monte Carlo Experiment II: Non-EPBR Data

Design

The second experiment follows the second experiment in Diamond & Sekhon (2006). The covariates are taken from the Dehejia and Wahba (1999) experimental sample of the LaLonde (1986) data (see the empirical application below and Diamond & Sekhon (2006) for details). The covariates are not ellipsoidally distributed and thus the EPBR conditions do not hold. Assuming a constant treatment effect of $1,000 the fictional earnings are a non-linear function of only two covariates:

$$Y = 1000D \; + \; .1\,exp[\,.7\,log(\texttt{re74} + .01)] \; + \; .7\,log(\texttt{re75} + .01)]) + \epsilon$$

where $\epsilon \sim N(0, .5)$, `re74` and `re75` are real earnings in 1974 and 1975 and D is the treatment indicator. The true propensity score is:

$$\pi_i = logit^{-1}[1 + .5.\mu + .01log(\texttt{age}^2) - .3log(\texttt{educ}^2) - .01log(\texttt{re74} + .01)^2 + .01log(\texttt{re75} + .01)^2]$$

[15] where the linear predictor μ is obtained from regressing the actual treatment indicator on $\texttt{age}^2$, $\texttt{educ}^2$, `black`, `hispanic`, `married`, `nodegree`, $\texttt{re74}^2$, $\texttt{re75}^2$, `u74`, and `u75` in the Dehejia Wahba sample. So the true propensity score is a combination of this logistic regression plus the extra variables specified in the equation above.

[15]Notice that this corrects a typo in the Diamond & Sekhon (2006) draft which omitted the logs for age and educ.

In Monte Carlo replications we use the following incorrect functional form to estimate the propensity score:

$$\begin{aligned}\hat{\mu} &= \alpha_0 + \alpha_1\texttt{age} + \alpha_2\texttt{educ} + \alpha_3\texttt{black} + \alpha_4\texttt{hispanic} + \\ &\quad \alpha_5\texttt{married} + \alpha_6\texttt{nodegree} + \alpha_7\texttt{re74} + \alpha_8\texttt{re75} + \alpha_9\texttt{u74} + \alpha_{10}\texttt{u75}\end{aligned}$$

We run 1000 simulations and report the mean estimate which indicates the bias (multiplied by 100) and the root mean squared error (MSE). We also report the average computing time per simulation measured in seconds.

Results

The results are displayed in table 3.2. Synthetic matching achieves the second lowest bias and the lowest MSE across all estimators. It is also much faster compared to Genetic Matching, which achieves the second lowest MSE in this experiment. The propensity score methods perform badly given the incorrect specification of the propensity score model. This indicates that in this real data situation where the EPRB conditions do not hold, synthetic matching retains good finite sample properties.

3.4.4 Monte Carlo Experiment III: Non-EPBR Data with Various Designs

Design

The third Monte Carlo experiment follows the design presented in Frölich (2007) who provides the most comprehensive investigation of the properties of propensity score adjustments to date. To make it more realistic we extend this design by considering a mixture of continuous and binary variables. We also examine additional factors such as the ratio of treated to controls and the degree of propensity score misspecification. The idea in this design is to mirror a range of typical scenarios that may be encountered in empirical settings in the social sciences.

We use six covariates X_j with $j \in (1, 2, ..., 6)$: X_1, X_2, and X_3 are multivariate normal with means zero, variances of $(2, 1, 1)$ and covariances of $(1, -1, -.5)$ respectively; X_4 is distributed uniform on $[-3, 3]$; X_5 is distributed χ_1^2; X_6 is Bernulli with $p = .5$. The treatment and control group are formed using:

$$D = \mathbf{1}[X_1 + 2X_2 - 2X_3 - X_4 - 0.5X_5 + X_6 + \epsilon > 0]$$

Notice that the covariates are weighted unequally as is reasonable in many empirical settings. We consider three designs for the error term ϵ which relate to different distributions for the true propensity score:

- Sample Design 1: $\epsilon \sim N(0, 30)$
- Sample Design 2: $\epsilon \sim N(0, 100)$
- Sample Design 3: $\epsilon \sim \chi_5^2$ and scaled to mean .5 and variance 67.6

Figure 3.1 visualizes the densities of the true propensity score in the three designs. The designs differ in the amount of separation between the treatment and control group. The first design shows the strongest separation and provides a fairly difficult case for matching. In the second design there is fairly weak separation and the estimators are expected to be more precise. The third design provides a middle ground as the variance lies between the first and the second design. However, the error term is leptokurtic so that the probit estimator which will be used to estimate the propensity score is misspecified.

We consider three sample sizes $n \in (300, 600, 1800)$ and also vary the ratio of control to treated units $r = n_0/n_1$ with $r \in (1, 2, 5)$ by sampling the specified numbers of treated and control units. For the estimators that rely on the estimated propensity score we use three different probit specifications with the following mean functions:

- PS Design 1: $\widehat{p(x)} = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 X_6$
- PS Design 2: $\widehat{p(x)} = \alpha_0 + \alpha_1 X_1^2 + \alpha_2 X_2^2 + \alpha_3 X_3 + \alpha_4 X_4^2 + \alpha_5 X_5^2 + \alpha_6 X_6$
- PS Design 3: $\widehat{p(x)} = \alpha_0 + \alpha_1 X_1 X_3 + \alpha_2 X_2^2 + \alpha_3 X_4 + \alpha_4 X_5 + \alpha_5 X_6$

These functions are designed to yield various degrees of misspecification of the propensity score model. For normal ϵ (Sample Designs 1 and 2) the first model is correct, the second model is slightly misspecified, and the third model is heavily misspecified. The correlations between the true and the estimated propensity scores are 1, .8, and .3 respectively. For non-normal ϵ (Sample Design 3) all three are misspecified, again with increasing levels of misspecification.

Finally, we consider three outcome designs:

- Outcome Design 1: $Y = X_1 + X_2 + X_3 - X_4 + X_5 + X_6 + \eta$
- Outcome Design 2: $Y = X_1 + X_2 + 0.2\,X_3\,X_4 - \sqrt{X_5} + \eta$
- Outcome Design 3: $Y = (X_1 + X_2 + X_5)^2 + \eta$

with $\eta \sim N(0,1)$. These regression functions are designed to mirror various degrees of non-linearity in the mapping of the covariates to the outcome. The true treatment effect is fixed at zero for all units. The different outcomes also exhibit different correlations with the true propensity score decreasing from .8, .54, to .16 from sample design 1 to 3 respectively. Again, we run 1000 simulations and report the mean estimate (bias) and the root mean squared error (MSE).

Results

The full results for $N = 300$, $N = 600$, and $N = 1500$ are shown in tables 3.3, 3.4, and 3.5 respectively.[16] The results are fairly similar across sample sizes except for the propensity score which improves in larger samples as expected. Figure 3.2 presents a graphical summary for the case of $N = 300$ and $r = 5$.

Overall, the results suggest that synthetic matching outperforms all other matching techniques in terms of MSE. This result is robust for all three sample designs, the three outcome specifications, the three ratios of controls to treated, and the three

[16] Notice that we omit Genetic matching for $N = 600$ and $N = 1500$ due to its excessive runtime compared to all other methods. There is no theoretical reason to expect Genetic matching to perform relatively better in larger samples so the $N = 300$ are representative of its general performance.

propensity score equations. The gains in MSE are often substantial. For example, in the most difficult case of sample design 1 (strong separation), $N = 300$, and the highly non-linear outcome design 3, the MSE from synthetic matching is about 2.6 times lower that of Genetic matching, 3.4 times lower than pair matching on a propensity score that is estimated with correctly specified probit regression, 3.9 times lower than Mahalanobis distance matching, and 4.6 times lower than weighting on estimated propensity score. As expected we find that weighting or matching on misspecified propensity scores (PS designs 2 and 3) results in much higher MSE even in large samples.

Synthetic matching also outperforms all other matching techniques in terms of bias, except in larger samples where matching and weighting on the correctly specified propensity score yields equally good bias performance as one would expect given that stochastic balancing improves. Yet, in these cases synthetic matching retains lower MSE even at a sample size of $N = 1500$. This demonstrates the efficiency gains in finite samples even when the propensity score is correctly estimated. In these cases the propensity score stochastically balances the covariates asymptotically but synthetic matching can improve on the balance in finite samples.

3.5 Application: The LaLonde Dataset

To illustrate the use of synthetic matching we also present an application to the LaLonde (1986) dataset, a canonical benchmark in the causal inference literature. The LaLonde data consists of an experimental and a non-experimental part. The first part is an experimental dataset that contains measures about subjects that participated in a randomized job training program, the National Supported Work Demonstration Program (NSW). We use this data to establish a benchmark estimate of the effect of the program. In the next step we replace the experimental control group with a control group drawn from non-experimental survey data where we measure the same covariates as in the experimental data. The challenge then is to conduct covariate adjustment in the non-experimental data to recover the results obtained from the randomized experiment. The extensive debate surrounding this dataset is reviewed in Diamond & Sekhon (2006).[17]

3.5.1 Data

We focus on the Dehejia and Wahba (DW) subset of the experimental Lalonde data which contains 185 treated and 260 control observations. The outcome of interest is post-intervention earnings from the year 1978. The pretreatment information includes previous earnings from the years 1974 and 1975 (`re74` and `re75` in US $), education (`educ` in years of schooling), age (`age` in years), ethnicity (binary indicators for `black` and `hispanic`), and marital status (binary indicator for `married`). All earnings data is zero for those who are unemployed in a particular year (binary indicators `u74` and `u75`).

We follow the majority of the previous literature and define the experimental target as the simple (unadjusted) difference in the means of 1978 earnings across treatment and control groups. By this metric the estimated average treatment effect for the treated is $1,794 with a .95 confidence interval from $551 to $3,038. We then replace the data from the experimental control group by non-experimental survey data from the DW subset of the Matched Current Population Survey-Social Security Administration file (CPS-1) (as in Diamond & Sekhon (2006)).

We conduct synthetic matching using the 10 raw variables, all their one-way interactions, and squared terms for the continuous variables `age` and `educ`. We omit squared terms for `re74` and `re75` and their interaction because due to their collinearity they are simply balanced by adjusting on the lower order terms.[18] We exclude nonsensical interactions such as between `nodegree` and `educ`, `hispan` and `black`, `u74` and `re74`, or `u75` and `re75`. Overall this results in 52 moments conditions.

[17] A selected list of papers that have used the Lalonde data to evaluate the performance of estimators includes (Dehejia & Wahba 1997, Dehejia & Wahba 1999, Dehejia & Wahba 2002, Dehejia 2005, Smith & Todd 2001, Smith & Todd 2005*a*, Smith & Todd 2005*b*)

[18] For example, their t-test p-values in the matched data are 0.76, 0.83, 0.99 respectively.

3.5.2 Balance Before Matching

Table 3.6 compares the distribution of the 52 covariates in the NSW experimental treatment group and the CPS-1 subsample using three conventional balance statistics, the standardized difference in means, the p-value for a difference of means tests, and the ratio of the variances between the two groups. Evidently, there are extremely stark differences between the two groups. For example, all mean differences are significantly different from zero at any conventional significance level, except the indicator for Hispanic ethnic background and very few of the interaction terms.[19] Not surprisingly, the unadjusted difference in means yields an estimate for the average treatment effect for the treated of $-8,506.50 with a standard error of 712.77, which is very far away from the experimental target answers of $\$1,794$.

3.5.3 Balance After Matching

Table 3.7 shows the balance results after synthetic matching. The covariate balance is now improved dramatically. All the means are almost exactly identical on all 52 characteristics. Almost all of the variance ratios are very close to the ideal value of one. By this metric, the degree of balance is much higher than in any previous study conducted on this dataset. Figure 3.3 also shows QQ plots for the four continuous variables age, education, and the two years of pretreatment earnings comparing the treated and control units in the matched (gray dots) and the unmatched data (black dots). Evidently, after synthetic matching the distributions are now much closer; the points cluster fairly closely around the 45 degree line.

Notice that some of the baseline variables are discrete and others contain point masses and skewed distributions which indicates that the conditions for EPBR do not hold and that it may be very difficult to find a propensity score specification via trial and error that would jointly balance all covariates. Figure 3.4 illustrates this graphically by comparing the standardized differences in means achieved by con-

[19]Notice that we use p-values here as a measure of balance, and not to conduct hypothesis tests in the conventional sense (see Imai et al. (2008)).

ventional propensity score weighting where the propensity score is estimated with a logistic regression in all 52 covariates and synthetic matching. While the propensity score weighting clearly moves us some way towards better balance, many covariates still remain fairly imbalanced with standardized bias greater than $|.1|$ (a table with full balance results is in the appendix table 3.9). Synthetic matching, in contrast, matches the means exactly which results in a standardized difference of zero for all covariate combinations.

3.5.4 Effect Estimates

Table 3.8 shows how the differences in balance translate into different estimates for the average treatment effect for the treated. The final treatment effect answer obtained from synthetic matching is \$ 1702 with a .95 confidence interval of \$ 275 to \$ 3128. This is answer is very close to the experimental target answer. The estimate is also slightly more efficient than that of Diamond & Sekhon (2006) who report a final estimate of 1734 with a .95 confidence band between -298 and 3766.

3.6 Conclusion

The goal of matching is to generate well-balanced samples, but commonly-used matching methods make it difficult for researchers to achieve high balance targets in many empirical settings. The fundamental problem with commonly-used matching methods is that they fail to focus on covariate balance directly, but rely on an intricate and often ineffective process of "manually" iterating between propensity score modeling, matching, and balance checking to search for a suitable matching solution. In the worst case, these techniques may increase bias for the subsequent estimation of treatment effects when balance improvements in some covariates are accompanied with decreased balance for other important covariates.

We introduce synthetic matching as a technique to create better balanced samples. The fundamental difference between synthetic matching and other commonly-used matching methods is that it directly focuses on covariate balance. The user specifies her desired balance level by imposing a potentially large set of moment conditions, and synthetic matching then derives a set of unit weights that exactly accommodate the balance constraints but retains as much information as possible. This obviates the need for continual balance checking and makes it easy for the user to reach highly balanced samples that are identical on a large set of moments. Through extensive Monte Carlo simulations and an empirical application to the LaLonde data, we show that these balance improvements translate into treatment effect estimates with lower approximation error and higher efficiency in finite samples compared to all other commonly-used matching methods.

While synthetic matching leads to better balance and greatly simplifies the use of matching for practitioners, it is important to notice that other problems that are commonly associated with matching methods (and covariate adjustment more generally) still apply. For example, synthetic matching provides no safeguard against bias from unmeasured confounders that are often a vexing problem in observational studies.

3.7 Tables for Chapter 3

Table 3.1: Results for First Monte Carlo Experiment

Estimator:	RAW	PSM	MD	PSMD	GM	PSW	SM
Design A: equal variances							
Bias	−60.19	−2.34	−9.66	−9.83	−11.36	−0.58	0.02
MSE	45.51	4.87	2.31	2.47	3.25	1.21	0.23
Bias / Bias SM	2746.19	106.78	440.73	448.53	518.36	26.31	1.00
MSE / MSE SM	195.78	20.94	9.93	10.61	13.99	5.22	1.00
Design B: unequal variances in treatment and control group							
Bias	−59.79	−10.01	−20.84	−23.62	−13.37	−25.64	−0.11
MSE	46.58	5.17	7.07	8.78	3.88	9.38	0.25
Bias / Bias SM	564.52	94.54	196.76	222.99	126.19	242.09	1.00
MSE / MSE SM	188.78	20.94	28.64	35.59	15.73	38.04	1.00
Design C: equal variances and including squared terms							
Bias	−63.24	−3.41	−10.10	−10.39	−11.11	−0.89	0.15
MSE	48.18	4.94	2.42	2.62	3.06	1.24	0.25
Bias / Bias SM	420.09	22.64	67.11	69.02	73.78	5.92	1.00
MSE / MSE SM	190.98	19.57	9.61	10.38	12.13	4.90	1.00

Note: Three independent normal covariates X drawn for 50 treated units with means .2 and for 150 control units with means 0. Outcome mapping is linear $Y = X\beta + \epsilon$ with $\beta = (1, 1, 1)'$ and $\epsilon \sim N(0, 1)$. The true treatment effect is zero for all units. 1000 simulations. *Design A*: Equal unit variances in both groups. Assumptions satisfy the conditions for EPBR in Rubin and Thomas (1992). *Design B*: Unequal variances: 1.5 in treatment and .5 in control group. *Design C*: Equal unit variances. Squared terms of X are included for all estimators, but not the outcome. This mirrors the situation of adjusting for additional irrelevant covariates. Raw: Difference of means; PSM: Propensity score matching; MD: Mahalanobis Distance Matching, PSMD: MD matching using PS and orthogonalized covariates; GM: Genetic Matching; PSW: weighting on PS; SM: synthetic matching. Matching is 1:1 pair matching. The propensity score is estimated with a linear logit in X.

Table 3.2: Results for Second Monte Carlo Experiment

	RAW	MD	GM	PS	PSMD	PSW	SM
Bias	−450	384	61	93	496	−183	−78
MSE	1632	690	518	1050	782	982	464
Time	0	0	23	0	0	0	0
Bias / Bias SM	5.8	4.9	0.8	1.2	6.4	2.3	1
MSE / MSE SM	3.5	1.5	1.1	2.3	1.7	2.1	1
Time / Time SM	0.0	1.0	1186.5	1.0	1.5	0.5	1

Note: $1,000 is the true effect for all units. The experiment follows the second experiment presented in Diamond & Sekhon (2006). The conditions do not satisfy EPBR and the mapping between the baseline covariates and the outcome is non-linear. The propensity score is misspecified.

Table 3.3: Results for Second Monte Carlo Experiment (N=300)

Sample Design 1: Strong Separation and Normal Errors

MSE	RAW	MD	GM	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	332	35	29	27	384	193	68	385	200	25	370	188	4
Ratio CtoT 1 Y2	502	24	15	14	163	532	71	193	442	26	161	537	5
Ratio CtoT 1 Y3	1196	1355	898	1186	1676	1835	2266	1995	2377	1590	1486	1633	346
Ratio CtoT 3 Y1	326	29	21	23	369	191	56	370	196	17	362	184	4
Ratio CtoT 3 Y2	495	20	10	11	155	523	57	175	445	18	157	528	4
Ratio CtoT 3 Y3	1269	1197	723	1108	1586	1849	2054	1777	2316	1074	1356	1666	291
Ratio CtoT 5 Y1	341	31	19	24	400	210	53	383	210	14	385	195	4
Ratio CtoT 5 Y2	512	20	12	13	165	550	52	172	473	15	166	547	5
Ratio CtoT 5 Y3	1471	1154	693	1184	1622	2041	2028	1892	2460	942	1378	1723	325
BIAS	**RAW**	**MD**	**GM**	**PSM1**	**PSM2**	**PSM3**	**PSMD1**	**PSMD2**	**PSMD3**	**PSW1**	**PSW2**	**PSW3**	**SM**
Ratio CtoT 1 Y1	179	52	42	13	189	133	79	192	138	13	188	135	1
Ratio CtoT 1 Y2	222	43	29	16	122	227	81	134	208	15	122	230	3
Ratio CtoT 1 Y3	304	350	215	125	359	404	461	428	472	98	347	389	102
Ratio CtoT 3 Y1	177	47	33	7	184	132	70	187	136	12	186	133	−1
Ratio CtoT 3 Y2	221	38	22	11	118	225	71	127	208	16	121	227	1
Ratio CtoT 3 Y3	302	322	206	97	332	400	434	398	461	102	334	389	85
Ratio CtoT 5 Y1	179	45	27	9	189	135	65	187	138	14	191	134	0
Ratio CtoT 5 Y2	223	35	19	10	119	227	65	122	211	17	123	229	2
Ratio CtoT 5 Y3	301	302	182	75	315	400	416	396	460	90	325	384	55

Sample Design 2: Weaker Separation and Normal Errors

MSE	RAW	MD	GM	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	148	15	8	12	137	91	28	146	97	7	129	83	2
Ratio CtoT 1 Y2	229	8	4	6	59	231	21	58	192	8	54	230	2
Ratio CtoT 1 Y3	654	661	364	590	655	994	1156	1016	1383	573	482	825	196
Ratio CtoT 3 Y1	151	14	7	12	138	89	24	140	99	5	130	82	2
Ratio CtoT 3 Y2	225	7	4	7	60	225	18	52	196	6	54	225	3
Ratio CtoT 3 Y3	777	660	360	575	673	1063	1114	984	1421	428	517	902	208
Ratio CtoT 5 Y1	162	16	7	19	159	103	26	154	104	5	144	88	3
Ratio CtoT 5 Y2	236	9	6	10	66	248	20	56	214	6	58	238	3
Ratio CtoT 5 Y3	963	642	349	822	872	1233	1080	1013	1507	482	527	943	288
BIAS	**RAW**	**MD**	**GM**	**PSM1**	**PSM2**	**PSM3**	**PSMD1**	**PSMD2**	**PSMD3**	**PSW1**	**PSW2**	**PSW3**	**SM**
Ratio CtoT 1 Y1	117	32	18	4	107	88	48	114	94	1	108	87	1
Ratio CtoT 1 Y2	149	19	9	5	69	147	40	70	134	2	68	148	2
Ratio CtoT 1 Y3	194	238	149	31	180	281	321	297	353	4	177	266	27
Ratio CtoT 3 Y1	117	29	14	3	106	86	43	110	94	3	108	87	−1
Ratio CtoT 3 Y2	147	16	6	4	68	144	36	64	135	3	68	146	−0
Ratio CtoT 3 Y3	209	231	140	47	177	281	309	285	353	39	185	273	33
Ratio CtoT 5 Y1	118	27	10	3	109	87	41	111	92	2	111	87	−0
Ratio CtoT 5 Y2	149	16	5	5	67	147	33	62	136	3	68	148	0
Ratio CtoT 5 Y3	198	210	119	17	165	276	285	270	341	18	174	263	5

Sample Design 3: Medium Separation and Leptokurtic Errors

MSE	RAW	MD	GM	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	226	26	20	23	251	147	47	254	159	14	246	144	3
Ratio CtoT 1 Y2	350	18	11	11	116	404	46	131	337	18	122	416	3
Ratio CtoT 1 Y3	1213	1174	757	1069	1143	1534	1983	1680	2024	775	1098	1315	374
Ratio CtoT 3 Y1	221	23	16	20	253	153	41	250	156	12	246	142	3
Ratio CtoT 3 Y2	343	16	9	11	114	402	38	118	341	17	122	408	3
Ratio CtoT 3 Y3	1212	1046	645	1050	1011	1521	1769	1465	1934	739	951	1281	354
Ratio CtoT 5 Y1	239	25	15	24	287	161	38	268	169	12	273	150	3
Ratio CtoT 5 Y2	355	16	10	12	125	421	34	118	363	15	127	428	4
Ratio CtoT 5 Y3	1563	1122	701	1328	1253	1799	1828	1562	2124	888	1061	1465	440
BIAS	**RAW**	**MD**	**GM**	**PSM1**	**PSM2**	**PSM3**	**PSMD1**	**PSMD2**	**PSMD3**	**PSW1**	**PSW2**	**PSW3**	**SM**
Ratio CtoT 1 Y1	146	45	35	12	150	116	65	154	123	26	152	117	0
Ratio CtoT 1 Y2	185	36	23	16	101	197	64	110	180	35	107	201	4
Ratio CtoT 1 Y3	306	325	217	123	270	362	430	390	434	201	295	343	145
Ratio CtoT 3 Y1	144	41	29	9	150	117	59	152	120	25	152	116	1
Ratio CtoT 3 Y2	183	32	20	15	98	195	57	102	180	34	106	199	4
Ratio CtoT 3 Y3	295	300	202	91	239	354	400	358	419	186	272	336	130
Ratio CtoT 5 Y1	147	39	23	9	157	116	53	153	122	22	159	117	1
Ratio CtoT 5 Y2	185	29	16	14	101	197	50	97	182	29	107	201	4
Ratio CtoT 5 Y3	309	291	186	85	221	364	387	347	417	187	268	342	122

Note: Results show MSE and Bias for each estimator. Six covariates with a mixture of continues, binary, and categorical variables. Experimental factors are: 3 sample designs as in figure 3.1 (sample design 1: strong separation and normal errors; sample design 2: weaker separation and normal errors; sample design 3: medium separation and leptokurtic errors), 3 outcome designs (Y1 linear: $Y1 = X_1 + X_2 + X_3 - X_4 + X_5 + X_6 + \eta$; Y2 somewhat non-linear $Y2 = X_1 + X_2 + 0.2\,X_3\,X_4 - \sqrt{X_5} + \eta$; Y3 highly non-linear: $Y3 = (X_1 + X_2 + X_5)^2 + \eta$), and 3 controls-to-treated ratios (Ratio CtoT 1, 3, and 5). Estimators are Raw: Difference of means; MD: Mahalanobis distance matching, GM: Genetic matching; PSM: Propensity score matching; PSMD: MD matching using PS and orthogonalized covariates; PSW: weighting on PS; SM: synthetic matching. All matching is 1:1 pair matching. We use three specifications (labeled with a 1, 2, or 3 postfix) for all propensity score based methods (PSM, PSW, PSMD). The first propensity score model is correct for sample designs 1 and 2, and slightly misspecified for sample design 3. Propensity score models 2 and 3 are increasing in misspecification (as measured by the linear correlation between the true and the estimated score). 1000 simulations for each scenario; the true treatment effect is zero.

Table 3.4: Results for Second Monte Carlo Experiment (N=600)

Sample Design 1: Strong Separation and Normal Errors

MSE	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	320	21	14	358	182	46	352	186	15	354	183	2
Ratio CtoT 1 Y2	497	15	7	153	524	49	164	449	19	152	529	2
Ratio CtoT 1 Y3	997	915	671	1326	1640	1764	1479	2026	742	1205	1526	172
Ratio CtoT 3 Y1	323	18	12	363	185	39	353	190	10	356	185	2
Ratio CtoT 3 Y2	499	13	5	153	525	41	155	459	12	156	532	2
Ratio CtoT 3 Y3	1041	807	599	1186	1675	1601	1357	2035	592	1123	1547	168
Ratio CtoT 5 Y1	324	17	14	368	186	33	352	190	8	360	182	2
Ratio CtoT 5 Y2	499	12	7	155	525	34	143	467	9	158	526	2
Ratio CtoT 5 Y3	1125	739	604	1133	1719	1431	1269	2022	627	1072	1589	186
BIAS	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	177	42	6	185	132	65	185	135	7	186	134	0
Ratio CtoT 1 Y2	222	35	10	121	227	68	126	211	8	121	229	2
Ratio CtoT 1 Y3	294	292	72	333	393	412	374	443	51	329	384	69
Ratio CtoT 3 Y1	178	38	7	187	133	60	185	136	11	186	134	0
Ratio CtoT 3 Y2	222	31	8	120	227	61	121	213	14	123	229	2
Ratio CtoT 3 Y3	296	272	63	311	393	390	355	441	76	314	384	61
Ratio CtoT 5 Y1	177	35	5	187	131	54	183	135	12	187	133	−1
Ratio CtoT 5 Y2	222	28	7	120	226	54	115	213	17	123	227	1
Ratio CtoT 5 Y3	294	253	58	284	389	362	335	432	96	302	384	50

Sample Design 2: Weaker Separation and Normal Errors

MSE	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	144	8	5	126	82	18	126	88	3	121	78	1
Ratio CtoT 1 Y2	223	4	3	53	222	13	44	191	3	50	221	1
Ratio CtoT 1 Y3	514	440	273	458	823	859	722	1143	219	372	767	95
Ratio CtoT 3 Y1	145	8	5	126	85	16	124	89	2	123	79	1
Ratio CtoT 3 Y2	224	4	3	51	227	12	42	199	3	50	224	1
Ratio CtoT 3 Y3	557	387	259	431	861	775	659	1149	200	354	783	103
Ratio CtoT 5 Y1	146	8	8	131	86	14	123	92	2	124	80	1
Ratio CtoT 5 Y2	225	4	4	54	230	10	38	204	3	51	225	2
Ratio CtoT 5 Y3	629	362	387	458	939	708	630	1138	215	330	786	130
BIAS	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	118	25	1	107	87	40	109	91	1	107	87	0
Ratio CtoT 1 Y2	148	15	2	68	147	33	63	136	1	68	147	1
Ratio CtoT 1 Y3	195	200	15	175	269	284	257	329	6	173	267	18
Ratio CtoT 3 Y1	117	23	2	107	87	37	107	91	3	108	87	0
Ratio CtoT 3 Y2	148	13	2	67	147	31	60	138	4	68	147	1
Ratio CtoT 3 Y3	199	184	18	161	271	266	241	326	22	165	267	14
Ratio CtoT 5 Y1	117	20	−1	107	86	32	104	91	2	107	86	−1
Ratio CtoT 5 Y2	147	10	1	67	147	26	54	138	4	67	147	−0
Ratio CtoT 5 Y3	198	171	6	146	270	246	226	316	22	154	261	3

Sample Design 3: Medium Separation and Leptokurtic Errors

MSE	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	218	16	12	239	138	32	233	150	10	234	139	1
Ratio CtoT 1 Y2	343	10	6	103	398	31	108	346	15	115	409	2
Ratio CtoT 1 Y3	1049	794	583	843	1346	1512	1219	1721	506	862	1210	234
Ratio CtoT 3 Y1	216	14	11	237	138	27	227	148	8	234	139	1
Ratio CtoT 3 Y2	341	9	5	104	396	27	99	348	13	116	405	2
Ratio CtoT 3 Y3	1069	680	577	761	1399	1352	1044	1644	512	789	1218	222
Ratio CtoT 5 Y1	224	14	13	249	144	25	232	152	7	246	141	2
Ratio CtoT 5 Y2	343	9	6	108	402	23	92	358	11	118	407	2
Ratio CtoT 5 Y3	1172	655	820	723	1462	1244	1006	1670	524	780	1233	238
BIAS	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	146	36	7	150	115	54	150	121	25	151	116	0
Ratio CtoT 1 Y2	184	29	11	98	198	53	101	184	35	105	201	4
Ratio CtoT 1 Y3	301	273	74	247	351	381	339	407	177	274	339	121
Ratio CtoT 3 Y1	145	32	6	149	114	50	147	119	24	151	116	0
Ratio CtoT 3 Y2	184	25	10	98	197	49	96	185	33	106	200	3
Ratio CtoT 3 Y3	302	250	63	217	357	358	309	395	181	263	339	117
Ratio CtoT 5 Y1	146	31	5	151	115	45	147	119	20	154	116	−0
Ratio CtoT 5 Y2	183	23	10	98	197	43	90	186	28	105	199	3
Ratio CtoT 5 Y3	297	237	42	195	351	335	292	388	167	245	333	103

Note: Results show MSE and Bias for each estimator. Six covariates with a mixture of continues, binary, and categorical variables. Experimental factors are: 3 sample designs as in figure 3.1 (sample design 1: strong separation and normal errors; sample design 2: weaker separation and normal errors; sample design 3: medium separation and leptokurtic errors), 3 outcome designs (Y1 linear: $Y1 = X_1 + X_2 + X_3 - X_4 + X_5 + X_6 + \eta$; Y2 somewhat non-linear $Y2 = X_1 + X_2 + 0.2\,X_3\,X_4 - \sqrt{X_5} + \eta$; Y3 highly non-linear: $Y3 = (X_1 + X_2 + X_5)^2 + \eta$), and 3 controls-to-treated ratios (Ratio CtoT 1, 3, and 5). Estimators are Raw: Difference of means; MD: Mahalanobis distance matching, GM: Genetic matching; PSM: Propensity score matching; PSMD: MD matching using PS and orthogonalized covariates; PSW: weighting on PS; SM: synthetic matching. All matching is 1:1 pair matching. We use three specifications (labeled with a 1, 2, or 3 postfix) for all propensity score based methods (PSM, PSW, PSMD). The first propensity score model is correct for sample designs 1 and 2, and slightly misspecified for sample design 3. Propensity score models 2 and 3 are increasing in misspecification (as measured by the linear correlation between the true and the estimated score). 1000 simulations for each scenario; the true treatment effect is zero.

Table 3.5: Results for Second Monte Carlo Experiment (N=1,500)

Sample Design 1: Strong Separation and Normal Errors

MSE	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	318	11	6	343	180	28	327	180	8	342	181	1
Ratio CtoT 1 Y2	496	8	2	147	526	30	134	467	10	146	530	1
Ratio CtoT 1 Y3	952	562	301	1097	1540	1262	1078	1863	488	1060	1498	95
Ratio CtoT 3 Y1	317	9	5	343	181	23	321	180	4	341	180	1
Ratio CtoT 3 Y2	495	6	2	143	522	24	123	468	6	149	524	1
Ratio CtoT 3 Y3	936	475	258	970	1547	1099	960	1833	280	977	1511	75
Ratio CtoT 5 Y1	318	9	5	343	180	20	318	181	4	348	181	1
Ratio CtoT 5 Y2	495	6	2	143	522	20	115	473	5	155	523	1
Ratio CtoT 5 Y3	988	420	263	837	1603	972	875	1833	263	910	1548	81
BIAS	**RAW**	**MD**	**PSM1**	**PSM2**	**PSM3**	**PSMD1**	**PSMD2**	**PSMD3**	**PSW1**	**PSW2**	**PSW3**	**SM**
Ratio CtoT 1 Y1	178	31	3	184	133	52	180	133	3	184	134	0
Ratio CtoT 1 Y2	222	26	6	120	229	54	115	216	5	120	230	2
Ratio CtoT 1 Y3	300	232	44	318	388	352	323	429	30	316	384	54
Ratio CtoT 3 Y1	177	28	3	184	133	47	178	133	9	184	134	−0
Ratio CtoT 3 Y2	222	22	4	119	228	48	109	216	12	121	228	1
Ratio CtoT 3 Y3	296	213	32	297	388	327	304	424	61	304	385	41
Ratio CtoT 5 Y1	177	26	2	183	133	42	176	133	12	185	134	0
Ratio CtoT 5 Y2	222	20	4	118	227	43	105	216	17	123	228	1
Ratio CtoT 5 Y3	297	197	37	266	391	305	287	421	91	291	387	39

Sample Design 2: Weaker Separation and Normal Errors

MSE	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	140	4	2	114	77	10	109	81	1	114	77	0
Ratio CtoT 1 Y2	221	2	1	47	219	7	32	197	2	47	220	0
Ratio CtoT 1 Y3	444	238	95	329	774	557	465	972	103	298	734	41
Ratio CtoT 3 Y1	140	4	2	115	78	9	104	82	1	115	77	0
Ratio CtoT 3 Y2	220	1	1	47	221	6	28	201	1	47	219	0
Ratio CtoT 3 Y3	448	206	104	289	754	495	409	954	80	277	731	42
Ratio CtoT 5 Y1	142	4	3	118	80	8	105	86	1	118	79	1
Ratio CtoT 5 Y2	221	2	1	48	222	5	28	205	1	49	220	1
Ratio CtoT 5 Y3	488	186	135	278	782	437	381	952	90	262	738	55
BIAS	**RAW**	**MD**	**PSM1**	**PSM2**	**PSM3**	**PSMD1**	**PSMD2**	**PSMD3**	**PSW1**	**PSW2**	**PSW3**	**SM**
Ratio CtoT 1 Y1	117	18	1	105	87	31	103	89	0	106	87	0
Ratio CtoT 1 Y2	148	10	1	67	147	25	55	139	1	67	148	0
Ratio CtoT 1 Y3	196	150	11	163	271	232	210	308	1	164	267	10
Ratio CtoT 3 Y1	117	16	−0	105	87	28	100	89	2	106	87	−0
Ratio CtoT 3 Y2	148	9	0	67	148	23	51	140	3	68	147	0
Ratio CtoT 3 Y3	196	139	6	152	267	218	195	304	18	158	266	8
Ratio CtoT 5 Y1	117	15	−0	106	87	25	100	91	3	107	87	−0
Ratio CtoT 5 Y2	148	7	0	67	147	20	49	141	4	68	147	−0
Ratio CtoT 5 Y3	196	127	0	136	266	200	184	299	26	150	264	2

Sample Design 3: Medium Separation and Leptokurtic Errors

MSE	RAW	MD	PSM1	PSM2	PSM3	PSMD1	PSMD2	PSMD3	PSW1	PSW2	PSW3	SM
Ratio CtoT 1 Y1	213	8	5	226	133	19	216	143	7	226	137	1
Ratio CtoT 1 Y2	339	5	3	98	390	19	87	354	13	112	403	1
Ratio CtoT 1 Y3	933	491	319	625	1264	1075	845	1468	379	745	1146	160
Ratio CtoT 3 Y1	217	7	5	230	134	17	212	144	6	231	137	0
Ratio CtoT 3 Y2	338	5	2	97	391	16	79	360	12	114	401	1
Ratio CtoT 3 Y3	967	426	291	557	1266	955	719	1450	386	700	1163	154
Ratio CtoT 5 Y1	216	7	6	229	135	15	208	144	5	235	136	1
Ratio CtoT 5 Y2	338	4	2	96	393	13	72	365	9	113	399	1
Ratio CtoT 5 Y3	1004	387	407	463	1307	850	640	1470	397	621	1189	166
BIAS	**RAW**	**MD**	**PSM1**	**PSM2**	**PSM3**	**PSMD1**	**PSMD2**	**PSMD3**	**PSW1**	**PSW2**	**PSW3**	**SM**
Ratio CtoT 1 Y1	145	27	4	149	114	43	146	119	24	149	116	−0
Ratio CtoT 1 Y2	184	21	7	98	197	42	92	188	34	105	200	3
Ratio CtoT 1 Y3	296	218	44	228	350	324	286	380	174	265	335	111
Ratio CtoT 3 Y1	146	25	3	150	115	40	144	119	23	151	117	−0
Ratio CtoT 3 Y2	183	19	7	97	197	39	87	189	33	106	200	3
Ratio CtoT 3 Y3	301	202	34	210	349	305	262	376	178	257	337	109
Ratio CtoT 5 Y1	146	24	4	149	114	36	142	118	20	152	116	−0
Ratio CtoT 5 Y2	183	18	6	96	197	35	82	190	29	105	198	3
Ratio CtoT 5 Y3	298	190	29	176	351	284	242	375	176	239	338	107

Note: Results show MSE and Bias for each estimator. Six covariates with a mixture of continues, binary, and categorical variables. Experimental factors are: 3 sample designs as in figure 3.1 (sample design 1: strong separation and normal errors; sample design 2: weaker separation and normal errors; sample design 3: medium separation and leptokurtic errors), 3 outcome designs (Y1 linear: $Y1 = X_1 + X_2 + X_3 - X_4 + X_5 + X_6 + \eta$; Y2 somewhat non-linear $Y2 = X_1 + X_2 + 0.2\,X_3\,X_4 - \sqrt{X_5} + \eta$; Y3 highly non-linear: $Y3 = (X_1 + X_2 + X_5)^2 + \eta$), and 3 controls-to-treated ratios (Ratio CtoT 1, 3, and 5). Estimators are Raw: Difference of means; MD: Mahalanobis distance matching, GM: Genetic matching; PSM: Propensity score matching; PSMD: MD matching using PS and orthogonalized covariates; PSW: weighting on PS; SM: synthetic matching. All matching is 1:1 pair matching. We use three specifications (labeled with a 1, 2, or 3 postfix) for all propensity score based methods (PSM, PSW, PSMD). The first propensity score model is correct for sample designs 1 and 2, and slightly misspecified for sample design 3. Propensity score models 2 and 3 are increasing in misspecification (as measured by the linear correlation between the true and the estimated score). 1000 simulations for each scenario; the true treatment effect is zero.

Table 3.6: Covariance Balance Before Matching: Treatment versus Control Group

covariate	mean.Tr	mean.Co	sdiff	var.ratio	p-value
age	25.82	33.23	−103.55	0.42	0.00
educ	10.35	12.03	−83.63	0.49	0.00
black	0.84	0.07	211.13	1.95	0.00
hispan	0.06	0.07	−5.30	0.84	0.47
married	0.19	0.71	−133.06	0.75	0.00
nodegree	0.71	0.30	90.44	1.00	0.00
re74	2095.57	14024.14	−244.11	0.26	0.00
re75	1532.06	13642.53	−376.19	0.12	0.00
u74	0.71	0.12	129.09	1.97	0.00
u75	0.60	0.11	99.89	2.48	0.00
age.age	717.39	1225.91	−117.92	0.30	0.00
educ.age	266.98	395.54	−139.00	0.35	0.00
educ.educ	111.06	152.90	−106.46	0.34	0.00
black.age	21.91	2.40	168.13	1.65	0.00
black.educ	8.70	0.81	187.46	1.99	0.00
hispan.age	1.36	2.38	−18.60	0.37	0.01
hispan.educ	0.58	0.73	−6.32	0.70	0.40
married.age	5.56	25.85	−169.81	0.42	0.00
married.educ	1.96	8.56	−158.05	0.49	0.00
married.black	0.16	0.05	30.49	3.05	0.00
married.hispan	0.02	0.05	−28.67	0.32	0.00
nodegree.age	17.97	10.09	60.08	0.59	0.00
nodegree.black	0.61	0.03	118.42	7.74	0.00
nodegree.hispan	0.05	0.04	4.67	1.25	0.53
nodegree.married	0.14	0.20	−15.98	0.77	0.03
re74.age	54074.04	509069.25	−344.89	0.10	0.00
re74.educ	22898.73	171241.56	−258.46	0.20	0.00
re74.black	1817.20	840.59	20.52	1.54	0.01
re74.hispan	151.40	893.68	−61.96	0.09	0.00
re74.married	760.63	11809.15	−301.78	0.12	0.00
re74.nodegree	1094.15	3432.61	−69.00	0.21	0.00
re75.age	41167.28	489047.95	−449.97	0.06	0.00
re75.educ	15880.57	167310.76	−445.70	0.07	0.00
re75.black	1257.04	804.32	14.73	0.69	0.05
re75.hispan	153.73	884.98	−72.07	0.07	0.00
re75.married	654.34	11366.04	−374.12	0.08	0.00
re75.nodegree	1134.96	3290.78	−72.80	0.18	0.00
u74.age	18.78	3.60	110.91	1.63	0.00
u74.educ	7.26	1.42	116.80	1.58	0.00
u74.black	0.60	0.01	120.04	23.63	0.00
u74.hispan	0.03	0.01	13.47	3.74	0.07
u74.married	0.11	0.06	17.19	1.83	0.02
u74.nodegree	0.52	0.05	94.73	5.28	0.00
u74.re75	307.44	175.27	11.61	0.69	0.12
u75.age	15.98	3.57	85.98	1.73	0.00
u75.educ	6.15	1.33	90.64	1.84	0.00
u75.black	0.52	0.01	101.29	22.19	0.00
u75.hispan	0.03	0.01	12.08	3.61	0.10
u75.married	0.09	0.06	7.70	1.31	0.30
u75.nodegree	0.43	0.04	79.78	7.08	0.00
u75.re74	43.85	203.65	−33.95	0.08	0.00
u75.u74	0.59	0.07	104.46	3.56	0.00

Note: mean.Tr: mean in treatment group; mean.Co: mean in control group; sdiff: standardized difference in means; var.ratio: ratio of variances (treatment/control group); p-value: p-value from difference of means test.

Table 3.7: Covariance Balance after Synthetic Matching: Treatment versus Matched Control Group

covariate	mean.Tr	mean.Co	sdiff	var.ratio	p-value
age	25.82	25.82	0	1.00	1
educ	10.35	10.35	0	1.00	1
black	0.84	0.84	0	1.00	1
hispan	0.06	0.06	0	1.00	1
married	0.19	0.19	0	1.00	1
nodegree	0.71	0.71	0	1.00	1
re74	2095.57	2095.57	0	1.14	1
re75	1532.06	1532.06	0	0.90	1
u74	0.71	0.71	0	1.00	1
u75	0.60	0.60	0	1.00	1
age.age	717.39	717.39	0	1.13	1
educ.age	266.98	266.98	0	1.02	1
educ.educ	111.06	111.06	0	1.01	1
black.age	21.91	21.91	0	1.00	1
black.educ	8.70	8.70	0	1.03	1
hispan.age	1.36	1.36	0	0.97	1
hispan.educ	0.58	0.58	0	0.94	1
married.age	5.56	5.56	0	1.04	1
married.educ	1.96	1.96	0	0.98	1
married.black	0.16	0.16	0	1.00	1
married.hispan	0.02	0.02	0	1.00	1
nodegree.age	17.97	17.97	0	0.98	1
nodegree.black	0.61	0.61	0	1.00	1
nodegree.hispan	0.05	0.05	0	1.00	1
nodegree.married	0.14	0.14	0	1.00	1
re74.age	54074.04	54074.04	0	1.14	1
re74.educ	22898.73	22898.73	0	1.23	1
re74.black	1817.20	1817.20	0	1.21	1
re74.hispan	151.40	151.40	0	0.79	1
re74.married	760.63	760.63	0	1.37	1
re74.nodegree	1094.15	1094.15	0	1.05	1
re75.age	41167.28	41167.28	0	1.00	1
re75.educ	15880.57	15880.57	0	0.93	1
re75.black	1257.04	1257.04	0	0.98	1
re75.hispan	153.73	153.73	0	0.65	1
re75.married	654.34	654.34	0	1.14	1
re75.nodegree	1134.96	1134.96	0	0.89	1
u74.age	18.78	18.78	0	0.99	1
u74.educ	7.26	7.26	0	1.01	1
u74.black	0.60	0.60	0	1.00	1
u74.hispan	0.03	0.03	0	1.00	1
u74.married	0.11	0.11	0	1.00	1
u74.nodegree	0.52	0.52	0	1.00	1
u74.re75	307.44	307.44	0	0.50	1
u75.age	15.98	15.98	0	0.99	1
u75.educ	6.15	6.15	0	1.01	1
u75.black	0.52	0.52	0	1.00	1
u75.hispan	0.03	0.03	0	1.00	1
u75.married	0.09	0.09	0	1.00	1
u75.nodegree	0.43	0.43	0	1.00	1
u75.re74	43.85	43.85	0	0.67	1
u75.u74	0.59	0.59	0	1.00	1

Note: mean.Tr: mean in treatment group; mean.Co: mean in control group; sdiff: standardized difference in means; var.ratio: ratio of variances (treatment/control group); p-value: p-value from difference of means test.

Table 3.8: Estimates for Average Treatment Effect for the Treated

	PATT	SE	T-value	.95 LB	.95 UB
Unadjusted	-8506	712	-11.93	-7109	9904
Regression Adjusted	1159	618	1.88	-52	2371
Experimental Target	1794	632	2.83	551	3038
Synthetic Matching	1702	727	2.33	275	3128

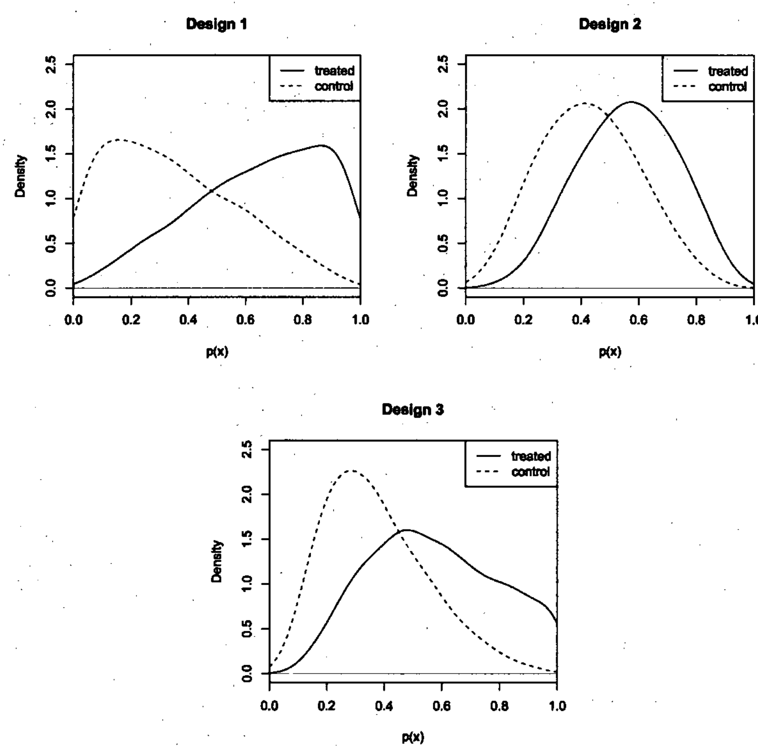
Table 3.9: Covariance Balance after Propensity Score Weighting: Treatment versus Matched Control Group

covariate	mean.Tr	mean.Co	sdiff	var.ratio	p-value
age	25.82	27.76	−27.13	0.92	0.01
educ	10.35	10.22	6.38	1.18	0.51
black	0.84	0.88	−8.76	1.21	0.36
hispan	0.06	0.04	6.80	1.35	0.47
married	0.19	0.16	8.26	1.16	0.39
nodegree	0.71	0.74	−6.77	1.07	0.49
re74	2095.57	1684.68	8.41	1.36	0.37
re75	1532.06	1307.01	6.99	1.07	0.48
u74	0.71	0.76	−11.41	1.13	0.24
u75	0.60	0.66	−12.18	1.07	0.22
age.age	717.39	826.13	−25.21	1.03	0.01
educ.age	266.98	282.20	−16.46	1.09	0.09
educ.educ	111.06	107.82	8.24	1.16	0.39
black.age	21.91	24.65	−23.63	1.00	0.02
black.educ	8.70	8.90	−4.72	1.25	0.62
hispan.age	1.36	0.96	7.11	1.36	0.45
hispan.educ	0.58	0.42	6.62	1.28	0.48
married.age	5.56	4.71	7.09	1.15	0.46
married.educ	1.96	1.58	9.16	1.20	0.34
married.black	0.16	0.13	7.42	1.17	0.44
married.hispan	0.02	0.01	2.27	1.21	0.81
nodegree.age	17.97	20.42	−18.74	0.89	0.07
nodegree.black	0.61	0.66	−10.99	1.07	0.26
nodegree.hispan	0.05	0.03	6.70	1.40	0.47
nodegree.married	0.14	0.11	7.49	1.19	0.44
re74.age	54074.04	43947.39	7.68	1.33	0.41
re74.educ	22898.73	18132.78	8.30	1.52	0.37
re74.black	1817.20	1451.45	7.69	1.46	0.40
re74.hispan	151.40	126.10	2.11	0.94	0.83
re74.married	760.63	608.57	4.15	1.73	0.64
re74.nodegree	1094.15	896.94	5.82	1.26	0.54
re75.age	41167.28	35832.82	5.36	1.14	0.58
re75.educ	15880.57	13332.83	7.50	1.14	0.44
re75.black	1257.04	1082.19	5.69	1.14	0.56
re75.hispan	153.73	122.50	3.08	0.88	0.76
re75.married	654.34	549.80	3.65	1.42	0.69
re75.nodegree	1134.96	983.57	5.11	1.06	0.60
u74.age	18.78	21.92	−22.99	0.95	0.02
u74.educ	7.26	7.71	−9.03	1.17	0.35
u74.black	0.60	0.68	−15.50	1.10	0.11
u74.hispan	0.03	0.02	6.61	1.55	0.47
u74.married	0.11	0.09	6.63	1.20	0.49
u74.nodegree	0.52	0.59	−12.21	1.03	0.22
u74.re75	307.44	310.15	−0.24	0.51	0.98
u75.age	15.98	19.22	−22.38	0.91	0.03
u75.educ	6.15	6.71	−10.44	1.11	0.28
u75.black	0.52	0.60	−15.74	1.04	0.11
u75.hispan	0.03	0.02	6.25	1.59	0.49
u75.married	0.09	0.07	4.86	1.17	0.61
u75.nodegree	0.43	0.50	−13.27	0.98	0.19
u75.re74	43.85	35.68	1.74	0.77	0.87
u75.u74	0.59	0.65	−12.53	1.07	0.20

Note: mean.Tr: mean in treatment group; mean.Co: mean in control group; sdiff: standardized difference in means; var.ratio: ratio of variances (treatment/control group); p-value: p-value from difference of means test.

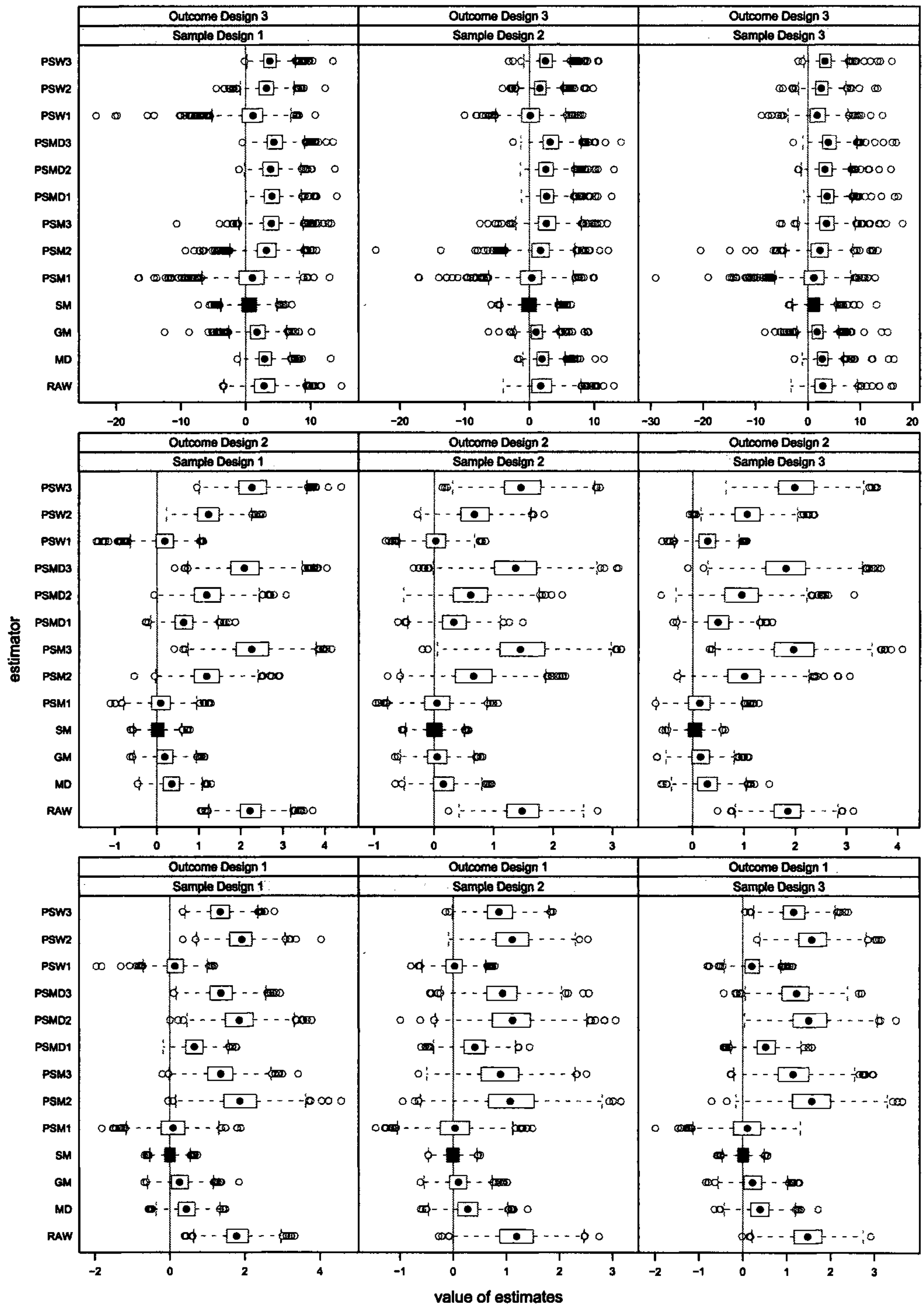
3.8 Figures for Chapter 3

Figure 3.1: Sample Designs for Second Monte Carlo Experiment: Density of True Propensity Score in Treatment and Control Group.



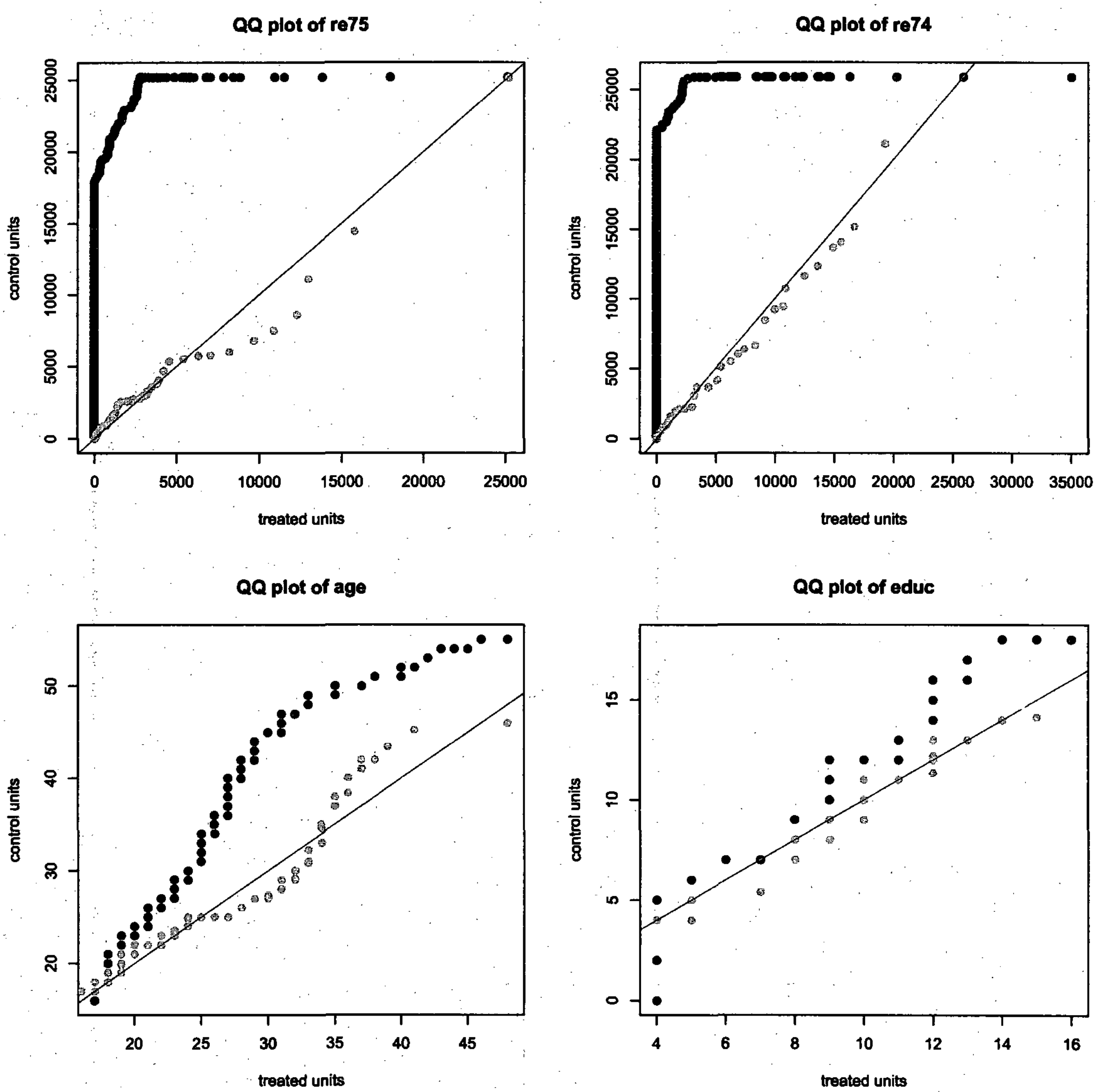
Note: Left graph refers to Sample Design 1 with $\epsilon \sim N(0, 30)$ (strong separation and normal errors); middle graph refers to Sample Design 2 with $\epsilon \sim N(0, 100)$ (weaker separation and normal errors); right graph refers to Sample Design 3 with $\epsilon \sim \chi^2_5$ and scaled to mean .5 and variance 67.6 (medium separation and leptokurtic errors).

Figure 3.2: Results for Second Monte Carlo Experiment ($r = 5$)



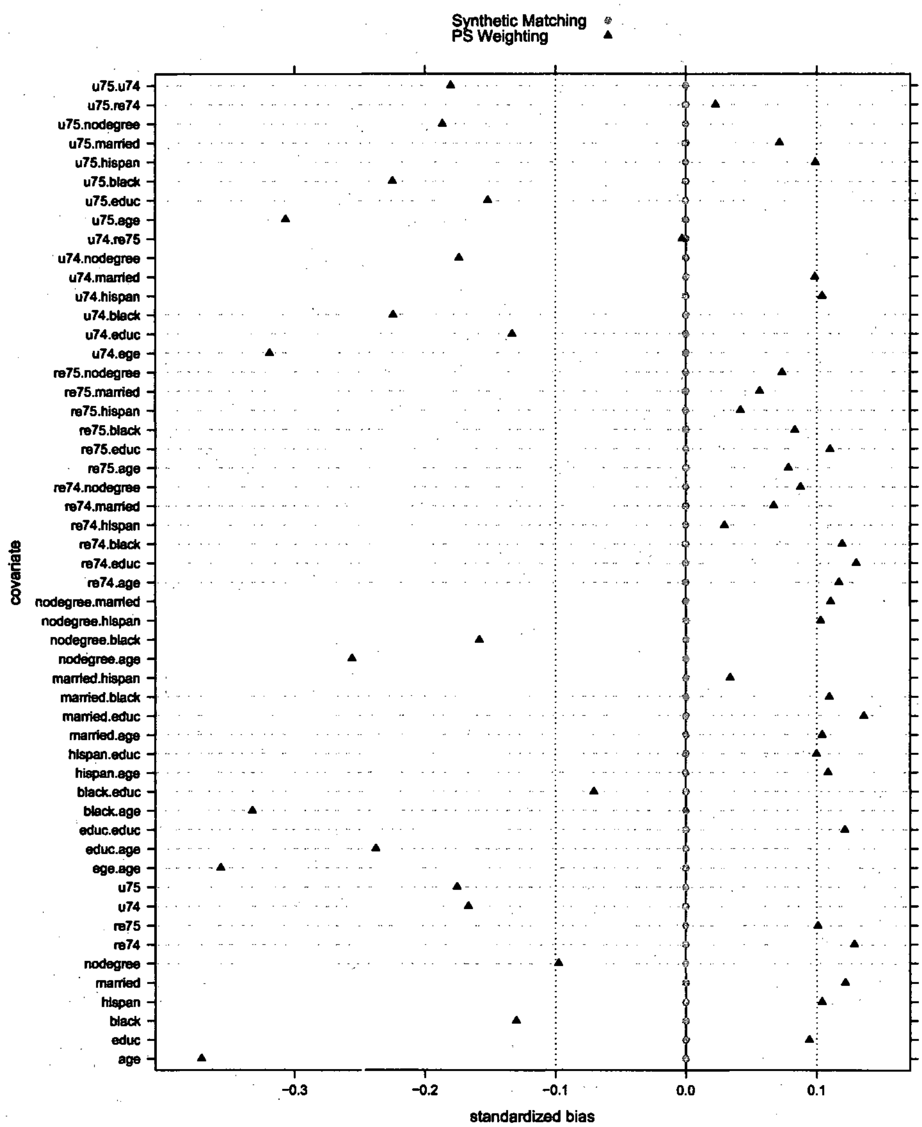
Note: Boxplots visualize the distribution of 1000 monte carlo estimates for the various estimators and scenarios. The true treatment effect is zero. Raw: Difference of means; MD: Mahalanobis distance matching; GM: Genetic matching; SM: synthetic matching; PSM1-3: Propensity score matching (with various degrees of misspecification in the PS model); PSMD1-3: MD matching using PS and orthogonalized covariates; PSW1-3: weighting on PS.

Figure 3.3: The Effect of Synthetic Matching on Covariate Balance: QQ plots of Continuous Covariates



Note: QQ plot of pretreatment earnings in 1975 and 1974, age, and education. The black dots represent empirical QQ estimates for the raw data. The gray dots represent QQ estimates for the matched data. The superimposed 45-degree line indicates identical distributions for the treatment and control group.

Figure 3.4: The Effect of Synthetic Matching on Covariate Balance: Synthetic Matching versus Propensity Score Weighting.



Note: Plot of covariate-by-covariate standardized bias after synthetic matching (gray circles) and propensity score weighting (black circles). The propensity score is estimated with a logistic regression in all 52 covariates. The standardized bias measures the difference in means between the treatment and control group (scaled by the standard deviation). Zero bias indicates identical means, dots to the right (left) of zero indicate a higher mean among the treatment (control) group.

Chapter 4

The Value of Political Power: Estimating Returns to Office in Post-War British Politics

4.1 Introduction

In October of 1989, Nigel Lawson resigned after six years as Chancellor of the Exchequer under Margaret Thatcher.[1] Four months later, while still a Member of Parliament, Lawson was named a non-executive director at Barclays Bank with a

[1]This paper is co-authored with Andy Eggers, Department of Government. This paper was awarded the 2009 Robert H. Durr Award from the Midwest Political Science Association (MPSA) for "the best paper applying quantitative methods to a substantive problem." A previous version was circulated under the title "MPs for Sale: Estimating the Returns to Office in Post-War British Politics." We thank Alberto Abadie, Jim Alt, Sebastian Bauhoff, Ryan Bubb, Jeff Frieden, Justin Grimmer, Adam Glynn, Torben Iversen, Mike Kellermann, Gary King, Roderick MacFarquhar, Clayton Nall, Riccardo Puglisi, Jim Robinson, Don Rubin, Ken Shepsle, Beth Simmons, Patrick Warren, Kevin Quinn, and seminar participants at Harvard, MIT, Penn State, and the NBER Political Economy Student Conference for helpful comments. For excellent research assistance we thank Matthew Hinds, Nami Sung, and Diana Zhang. We would especially like to thank Jim Snyder who directly inspired this project. The usual disclaimer applies.

salary of 100,000 GBP – roughly four times his MP pay. The afternoon the appointment was announced, Barclays' market value rose by nearly 90 million pounds (Hollingsworth 1991, pg. 150).

Anecdotes like this suggest that political connections can be of great value to private firms. In a number of recent papers, economists have begun to systematically examine this value in a variety of settings. Firms with personal and or financial connections to politicians enjoy higher stock valuations in Indonesia (Fisman 2001), the United States (Goldman, Rocholl & So 2008*a*, Jayachandran 2006, Roberts 1990), Malaysia (Johnson & Mitton 2003), and Nazi Germany (Ferguson & Voth 2008). In the United States politically connected firms are more likely to secure procurement contracts (Goldman, Rocholl & So 2008*b*) and in Pakistan they are able to draw more favorable loans from government banks (Khwaja & Mian 2005). Faccio (2006) shows that the benefits of political connections are larger in countries with higher corruption scores.

In this paper, we approach the market for political favors in the UK from the opposite perspective. Where others have focused on the benefits companies like Barclays obtain through connections to powerful politicians, we analyze the benefits politicians like Lawson obtain on the basis of their political power. If firms buy political favors, and if they do so in part by providing employment, gifts, or bribes to politicians, then politicians can be expected to benefit financially from office just as firms do from connections to officeholders. We attempt to measure this benefit by examining the effect of serving in Parliament on the estates of British politicians who entered the House of Commons between 1950 and 1970 and have since died.

Measuring the value of political power is difficult in part because detailed data on politicians' personal finances is generally not available. Even where it is, as in the US Congress since the early 1990s, we generally do not have good data about income or wealth after the member leaves office, when the financial value of political power may be realized (Diermeier, Keane & Merlo 2005). Even if we knew a given MP's income from all sources over the course of his life, it would still be difficult to determine what portion of those payments were a result of his political power. MPs are not randomly selected from the population (which is unfortunate for researchers but arguably beneficial for citizens), so a comparison of MPs' income or wealth with that of a peer group outside of politics is likely to reflect factors that led MPs to gain political office as well as the value of political office itself.

Our strategy for addressing these problems is to compare the wealth (at death) of MPs with that of politicians who ran for Parliament unsuccessfully. Voting, not randomization, decides which candidates win elections; we address the resulting selection problem in two ways. First, we employ conventional methods of covariate adjustment (matching and regression) to control for imbalances in key candidate-level confounding factors recorded in our dataset, including age, occupation, schools and universities attended, and titles of nobility. Second, we employ a regression discontinuity design (Thistlethwaite & Campbell 1960, Lee 2008), exploiting the quasi-random assignment of office in very close races to estimate the effect of office on wealth. Our estimation strategies yield the same basic result: serving in Parliament was quite lucrative for MPs from the Conservative Party but not for MPs from the rival Labour Party. Conservative MPs died almost twice as wealthy as similar Conservatives who unsuccessfully ran for Parliament; no such difference is evident among Labour politicians.[2]

Our identification strategy and rich set of covariates make us quite confident that the difference in wealth we observe between winning and losing candidates is due to serving in Parliament itself (as opposed to background differences between successful and unsuccessful politicians), but estimating that effect alone does not tell us *how* serving in office increased wealth for Conservative politicians. Serving in political office could affect one's wealth at death through many channels, including official perquisites (the office could provide a salary and in-kind payment different from what one could earn in the private sector), lifestyle changes (political culture could shape one's consumption patterns or bequest motive), and health (the stress or glory of being in Parliament might affect how long one accumulates and depletes savings). Our investigations suggest that these pathways do not account for the wealth gains

[2] As discussed below, our estimate measures the effect of power on bequest size; some consideration is required to translate that effect into the effect on earnings.

we observe among Conservative politicians. The official perquisites of office were modest in the period we examine, particularly compared to salaries in the occupations that Conservative candidates typically held before standing for office. We know of no particular lifestyle changes made by Conservative MPs that would substantially affect their personal finances or bequests.[3] Our analysis also reveals no effect of winning office on longevity.

We suggest that office was lucrative for Conservative politicians because it endowed them with political connections and knowledge which could be put to personal financial advantage. We show that winning office more than tripled the rate of corporate non-executive directorships obtained among Conservative politicians. Back-of-the-envelope calculations suggest that this difference in the number of directorships alone can account for a sizable portion of the wealth differential between MPs and unsuccessful candidates from the Conservative Party. These directorships and other forms of politically linked consultancies were lucrative for Conservative politicians, but also valuable for the firms that employed them because of the political knowledge and connections possessed by sitting and or former legislators. This finding complements evidence from several other studies that have shown that political connections add value to firms. (For example, in the U.S. Goldman et al. (2008*a*) finds that a company experiences a positive abnormal return following the announcement of a board nomination of a politically connected individual). We also provide some suggestive evidence that the ability to translate political knowledge into personal wealth is stronger for MPs with better access to political power. Conservative MPs who entered office in years when their party held power appear to have gained more than those who entered when it did not. Moreover, we find that among Conservative MPs an additional year in office is associated with a roughly 3 percent increase in wealth indicating that more political connections and knowledge is more valuable.

We argue that the larger benefit enjoyed by Conservative MPs was due in part to differences in the way the parties were financed and organized. The Labour Party was funded and dominated by a handful of trade unions that used their influence to

[3] A possible exception is that MPs were probably more likely to invest in London real estate.

secure the exclusive loyalty of a large proportion of Labour MPs. The Conservative Party by contrast gathered its financial support from diffuse contributors and had no dominant constituency, leaving MPs relatively free to forge relationships with numerous outside firms that competed for their legislative services. MPs from both parties thus explicitly provided services to outside interests, but the trade unions shaped Labour Party institutions such that they could acquire those services without making payments directly to the MPs.

Our paper is among the first to provide direct empirical estimates of the financial rewards of political office. It is closely related to Querubin & Snyder (2008), who use census data to assess whether members of the US Congress in the 19th century enjoyed faster wealth growth than unsuccessful Congressional candidates. Our estimates speak to the "career concerns" literature in political science, including work on candidate recruitment (Schlesinger 1966, Rohde 1979, Fiorina 1994, Osborne & Slivinski 1996, Besley & Coate 1997) and candidate retirement (Groseclose & Krehbiel 1994, Hall & van Houweling 1995, Diermeier et al. 2005, Keane & Merlo 2007). The monetary benefit of officeholding also appears as an important parameter in numerous recent political economy models that examine the selection and behavior of politicians (Caselli & Morelli 2004, Messner & Polborn 2004, Besley 2005, Besley 2006, Dal Bó, Dal Bó & Di Tella 2006, Mattozzi & Merlo 2007). There is no consensus in the theoretical literature on the relationship between the financial rewards of political office and the quality of policymaking; empirical work is only beginning to assess the evidence, examining the relationship between legislator quality and outside income (Gagliarducci, Nannicini & Naticchioni 2008) or official salary (Ferraz & Finan 2008). Our analysis can provide only suggestive evidence of the relationship between financial rewards and MP quality, but it furnishes the first estimates of the total financial rewards of attaining legislative office, demonstrates that non-salary benefits are a considerable part of those rewards, and shows that those rewards can vary widely by party.

4.2 The Value of a Parliamentary Seat in Context

Before embarking on our empirical analysis of the financial benefit of winning a seat in the House of Commons, it is worth illuminating the context surrounding MPs' finances. No study has previously attempted to empirically determine the total financial rewards of serving in Parliament, but there has been considerable controversy and discussion on the financial lives of MPs that allows us to form some expectations.

MPs earn salaries that are considered modest relative to their counterparts in other countries and in comparable professions within Britain (Judge 1984, Baimbridge & Darcy 1999), but there is a widespread public perception that some MPs use office to enrich themselves by other means. A Gallup poll in 1985 found that 48 percent of respondents thought that "most MPs make a lot of money by using public office improperly;" by 1994, when scandals surrounding Parliamentary bribes had become a prominent political issue, the proportion of respondents answering in the affirmative had risen to 64 percent and over 80 percent thought it improper for MPs to accept payment for advice about parliamentary matters (which is in fact a common practice in Parliament) (Norton 2003, pg. 367).

While outright bribery has occasionally been the focus of some attention (particularly in the "cash-for-questions" scandal of the mid-1990's, most public scrutiny has focused on the practice of MPs taking on outside employment while in office. As in most other parliaments, members of the British House of Commons are permitted to take on a variety of outside work while serving in office. Throughout the period since World War II, it has been common for MPs to serve on corporate boards, act as "parliamentary consultants" for firms or industry groups, and draw stipends from trade unions. While the practice of MPs simultaneously holding outside jobs is consistent with the concept of parliaments as citizens' assemblies, it has long been recognized that these outside arrangements might conflict with MPs' duties to serve the public interest and their constituencies. A number of exposés (Stewart 1958, Noel-Baker 1961, Finer 1962, Roth 1965, Judge 1984, Doig 1984, Hollingsworth 1991) highlighted these conflicts, often focusing on Conservative MPs, who were reportedly more likely to access lucrative outside employment due to their political connections and knowlegde. Debates surrounding members' salaries and outside interests, taken up both in Parliament and in the broader public sphere, presaged recent formal models on the issue of legislative compensation, e.g. Gagliarducci et al. (2008). Defenders of MPs' outside interests argued that members gained policy-relevant knowledge from their outside work and that banning parliamentary consultancies and directorships would drive the best MPs out of politics, while those advocating restrictions claimed that limiting outside employment would reduce conflicts and encourage sitting MPs to focus on their legislative work.

The Commons has addressed the potential conflict between legislative duties and outside interests by forbidding ministers from taking outside work and, since 1975, requiring other members to disclose the names of their outside employers from which they receive benefits which might reasonably be thought by others to influence their actions, speeches or votes in Parliament, or actions taken in the capacity of a MP. In 1996, following a scandal in which members were caught accepting payments for raising issues in Parliament, MPs were required to report amounts received from outside employment and expressly forbidden from carrying out "paid advocacy," but their right to take on work as consultants and directors while in office (and any work whatsoever afterward) was protected. This approach may seem lax from the perspective of the present-day US Congress, whose members are prohibited from taking on almost all outside employment, face strict caps on earned income, gifts, and travel, and are prohibited from taking lobbying employment during a "cooling off period" after leaving Congress.[4] Compared to other legislatures internationally, though, the UK's regulations on conflict of interest are quite typical (Faccio 2006).[5] What is unusual is the closeness of connections between British MPs and British

[4]Committee of Standards of Official Conduct, *House Ethics Manual*, 2008 edition.

[5]That regulations on British MPs are fairly typical is further confirmed by a 1999 report (Whaley 1999) surveying codes of conduct, disclosure rules, and employment restrictions in twenty countries of various levels of economic development. While a comparable survey of regulations in earlier periods has not been conducted, it is worth noting that there was little difference in the regulation of members' outside interests between Britain and the US until the late 1970s. Senators could serve on corporate boards until 1977, and members of the House as recently as 1990; a cap on outside earned income was first introduced in the House in 1977 and the Senate in 1990. See Susan F. Rasky, "Plan to Ban Fees Spurs Lawmakers," *The New York Times*, February 1, 1989.

industry: Faccio (2006) estimates that 39% of British firms (by market capitalization) have politicians in the executive ranks or as major shareholders, making the UK the third most connected country in her sample, behind only Russia and Thailand.[6] To provide a longer view of the extent of connections between sitting MPs and business in the UK, we recorded the outside interests reported by MPs for 1975 (the first year disclosure was required), 1990, and 2007. Figure 4.1 depicts the proportion of MPs, by party, who reported outside employment as directors, journalists, or consultants, as well the proportion of MPs who reported other employment (ie. unrelated to MP work), a union sponsorship, or significant shareholdings.[7] The plots indicate that a considerable proportion of MPs had outside engagements but, as might be expected, there were stark differences in the types of engagements undertaken by Conservative and Labour MPs. Around half of Conservative MPs sat on corporate boards at

[6]Faccio labels a firm as politically connected if an MP or government minister is either a top officer or a large shareholder as of 2001. Her estimate may overstate the extent of connections in the UK in comparison to other countries, since many of the connections she observes involve members of the House of Lords, a largely ceremonial body with no counterpart in most countries in her survey. (The "Register of Lords' Interests" confirms that peers are highly connected to business; see e.g. Jo Dillon, "One in three peers has seat in boardroom," *The Independent*, July 28, 2002.) Still, even if half of the connections she records are attributed to the House of Lords and thrown out, the UK remains among the top five most connected countries in the survey.

[7]We used editions of the Register of Member Interests published on 1 November 1975, 8 January 1990, and 26 March 2007. Details on each type of income, and our approach to recording it, are as follows: Directorships include only remunerated directorships. Consultancies include all remunerated consulting activities classified as parliamentary affairs advisor, economic advisor, liaison officer, public affairs consultant, parliamentary consultant, management consultant or advisor for firms when in connection to MP work, public relations consultant, public relations agents, members of parliamentary panels. Lloyd's underwriter are also included. We excluded all consulting declared as unremunerated, charitable, or obviously unrelated to commercial lobbying (eg. council work). We included consultancy work for trade union related groups. For 2007, we also included speech engagements that are clearly connected to consulting work. Journalism includes any type of remunerated journalistic activity such as broadcasting, TV appearances, newspaper, occasional journalism, novelists, documentaries, and scholarly articles, work as editor for the house magazine, and (especially in 2007) also book contracts. We excluded unremunerated journalistic activities and activities where fees are reported to be transferred to charities. Union sponsorship includes campaign support as well as continual sponsorship of sitting MPs. Employment includes regular employment that is declared as unrelated to MP work, such as work as a barrister at law, a partner in a law firm, medical practitioner, farmer, or family business, etc. We excluded work that is declared as infrequent (such as occasional work as Queen's Council). MPs are required to register shareholdings for any public or private company in which they hold more than 15 percent of the issued share capital or shares worth more than 100 percent of the official MP salary (for example 60,675 GBP in 2007).

each point examined, and around half reported employment as a "parliamentary consultant." Labour MPs were much less likely to hold either kind of position but, up until the 1990s, were very likely to be sponsored by a trade union. (The Labour Party ended union sponsorships in 1996 in part to sharpen its attacks on Conservatives' outside financial dealings.[8]) Plenty of anecdotal evidence suggests that the rough pattern of outside interests revealed by official disclosure starting in 1975 extends back well into the 1950s and 1960s.[9]

To this point we have considered outside employment in which MPs have engaged while in office, but some of the financial rewards of holding office probably come after an MP retires from politics (whether because payments for political services are delayed until they can be hidden or because the MP continues to provide political services). A distinct advantage of our research design (which uses probate values as the outcome) is that it should measure rewards MPs collect during their entire lives after winning office, including after they retire from politics. Because former MPs are not subject to disclosure requirements, we know far less about the employment opportunities they enjoyed after leaving office than before. For comparison, Diermeier et al. (2005) conclude based on a survey former Congressmen's first job after leaving office that Congressional experience confers a considerable boost in earning power.

[8] James Blitz, "Labour poised to end trade union sponsorship of MPs," *Financial Times*, February 28, 1996.

[9] Already in 1896 the Economist complained that "Notoriously, men are often placed on boards of directorship simply and solely because they are Members of Parliament and are, therefore, believed to be able to exercise unusual influence." *The Economist*, April 18, 1896. A sharp increase in the MP-as-lobbyist pattern occurred after World War II (see Stewart (1958) and Beer (1956) for early studies). In 1950 the Attlee Commission (convened to investigate outside interests and lobbying in the House of Commons) concluded that commercial lobbyists were "few in number," but by 1962, Finer notes a rising "army" of professional lobbyists and MPs under contract, noting that "Parliament is not 'above' the battle between associations and counter-associations; it is the cockpit" (Finer (1962, pg. 43), also see Stewart (1958) and Harrison (1960) for evidence on sponsored MPs in the 1950s and 1960s). In 1961, Labour MP Frances Noel-Baker estimated that the number of MPs employed by advertising and public relations firms had risen from 18 in 1958 to 27 in 1961 (Noel-Baker 1961), and Hollingsworth (1991, pg. 113) put this number at at least fifty in 1965. The *Business Background of MPs*, periodically published by journalist Andrew Roth beginning in 1957, confirms that the disproportionate involvement of Conservatives in consulting, directorships, and public relations was consistent throughout the careers of the MPs in our sample (Roth 1957). Similarly, Muller (1977) shows that between 1945 and 1975 over 30% of all Labour candidates and over 40% of all Labour MPs were directly sponsored by the unions.

The controversy surrounding lobbying income collected by former Senate Democratic Leader Tom Daschle after leaving the Senate in 2004 provides further anecdotal evidence from the U.S. that former legislators can profit from their political experience.[10]

4.3 The Wealth of Candidates to the House of Commons

4.3.1 Data and Estimation Sample

Our research design assesses the financial benefits of political office by comparing the wealth of MPs with that of unsuccessful candidates. In this section we describe the process by which we collected wealth data, along with relevant covariates, for a sample of winning and losing candidates to the British House of Commons.

As a measure of wealth, we focus on politicians' probate values, a legal record of the size of an individual's estate at the time of death.[11] Probate values are widely used as a measure of wealth by economic historians[12] and provide the basis for official statistics on the distribution of wealth even today.[13] Over 90% of UK citizens leave a

[10]Ceci Connolly, "Daschle Pays $100k in Back Taxes Over Car Travel," *The Washington Post*, January 30, 2009.

[11]In the UK, a probate is needed in order for a deceased person's representative to administer the assets of the estate. A probate is normally filed for all estates containing real property and/or a single class of asset worth 5,000 GBP or more. By law, the estate includes the value of all assets and monies at the time of death, after debts and expenses have been deducted, plus any gifts exceeding 3,000 GBP that have been made within the previous seven years and the value of any trust from which the deceased has received an income. Jointly held property is also exempt, with certain restrictions. At the time of writing, a 40% inheritance tax is applied to the estate, with the first 300,000 GBP exempt. Tax avoidance may affect the reported wealth but this effect is mitigated by the fact that gifts given within seven years of death are taxable.

[12]See Owens, Green, Bailey & Kay (2006) for an application, discussion, and many citations.

[13]In a recent review comparing methods of estimating the wealth distribution, HM Revenue & Customs (HMRC) concluded that the approach based on probate values remains "the best available means," surpassing alternate approaches based on investment income and direct household surveys (HMRC 2007, pg. 3).

probate record (the exceptions being mostly indigent people) and the probate values for residents of England and Wales since 1858 are available in a single archive in London that allows to collect the probate value for a person with a known name and date of death.

Since the biographies of MPs are typically listed in encyclopedias and official publications, the names and dates of death of successful candidates are easy to acquire. The primary difficulty is in finding the date of death of losing candidates, who for the most part leave a very scant historical trace. Fortunately, starting in the late 19th century the *Times of London* published brief biographies of every parliamentary candidate (winning and losing) standing for the House of Commons in each election. Since the candidate biographies are published at the time of the election, they do not of course provide the date of death. Still, the details provided by the biographies - in particular, the full candidate name along with the year and sometimes month of birth - are sufficient to locate many candidates in public death record archives. We used an online genealogy database[14] that indexed all death records filed since 1984 by year and month of birth, which made it quite straightforward to find the date of death for a candidate using the information in the *Times* biographies.[15] An additional benefit of the *Times* biographies is that they include information on the education, occupation, and sometimes family background of the candidates, characteristics which are likely to be correlated with the candidates' ability and wealth at the time they ran for office.

We therefore digitized the *Times Guide to the House of Commons* for each of the seven general elections between 1950 and 1970[16] and extracted key biographical and electoral information for every candidate (some 5,729 individuals). For each

[14] `www.thegenealogist.co.uk`

[15] Death records before 1984 are also available from this and other archives, but only as image files and not indexed by date of birth. This makes it much more time consuming to find earlier deaths, which led us to restrict our search to deaths since 1984.

[16] We chose the time period to maximize the number of candidates for whom we could find probate values. *The Times Guide to the House of Commons* did not provide candidates' years of birth before its 1950 edition, which sets the lower bound on our search range. We stopped collecting data after the 1970 election because candidates by then were young enough that a relatively small proportion would have died by now.

candidate, we record the full name, date of birth (year and, if available, month), education (both secondary and university), and occupation, as well as an indicator for whether he or she has a title of nobility. We then used the genealogy database to search for the date of death of 2,904 relatively competitive candidates, which at this stage we define as candidates who, not having previously won an election, either won or lost by fewer than 10,000 votes in a general election between 1950 and 1970. This restriction was intended to exclude incumbents, unbeatable candidates, and non-contenders for whom the implicit counterfactual is not well-defined. We found near-certain matches for 665 candidates; we were unable to find a record in cases where the candidate had not yet died, died before 1984 (the start of the death record database), or produced so many matching death records (because of a common name) that we were not able to identify the right one with sufficient certainty. In order to ensure the comparability of our winning and losing samples we ignored public information about winners' death dates and searched for the date of death in the same way for both MPs and losing candidates. This results in some known Type I and Type II errors in the sample of winners, but reduces the possibility that an observed difference in wealth between the two groups could be due to measurement error.[17]

With the 665 death records we obtained, we were then able to find probate values for 561 candidates in the probate calendar stored at First Avenue House in London.[18]

[17] To develop a protocol for finding death records given names and dates of birth, we created a sample of public figures (scientists, authors, athletes, etc.) whose death dates are publicly available from the *Oxford Dictionary of National Biography* and other sources and whose years of birth match the distribution in our sample of parliamentary candidates. We then searched the genealogy database for the death dates of these figures using only the last name and year/month of birth. For most names this search retrieves several possible matches, even in cases where the individual is not yet dead or died before the database's start year. We employed the random forests algorithm (Breiman 2001) to optimally identify correct matches using information about closeness of the name match and raw name frequency. Cross-validation indicated that we could achieve a Type I error rate of around 5%. Once we obtained death dates for our sample of parliamentary candidates using this algorithm, we checked our collected death dates against the true death dates for successful candidates (which are easily available from public records) and confirmed that we indeed had an error rate of 5.2%.

[18]The few missing probates were mostly due to common names. Probates are listed under the quarter in which they are registered, which might be as much as a year after the date when the death was registered, and entries in the probate calendar do not list birth dates (unlike death records). As a result, there might be several possible probate records listed in the year or so following the death of a candidate with a common name, making it impossible to tell which one is the correct estate.

We then exclude from our estimation sample 67 candidates who were from minor parties (36 Liberals and 31 from regional parties) and a further 67 candidates who were found to have served before 1950, which leaves us with 427 candidates overall. Of these, 165 candidates are "competitive winners" in the sense that they entered Parliament in a race they won by fewer than 10,000 votes; the remaining 262 candidates are "competitive losers" in the sense that at some point they came within 10,000 votes of winning.[19] The candidates in our estimation sample are spread quite evenly geographically across Britain (with candidates appearing in 383 out of 658 possible constituencies between 1950 and 1970) and temporally across our period (with about 60 candidates making their debut in each of the seven elections between 1950 and 1970). As far as we know, our database is unique in the richness of the background information and electoral results it provides about both winning and losing candidates over several elections. With Querubin & Snyder (2008), we are also among the first to collect direct measures of candidate wealth.

4.3.2 Wealth Distributions

Table 4.1 provides descriptive statistics on the distribution of wealth at the time of death for candidates in our sample. To make the comparison meaningful, we converted the gross value of the estate into real 2007 British Pounds (GBP) using the Consumer Price Index from the Office for National Statistics. We find that gross wealth at death varies widely across candidates ranging from 4,597 GBP for the poorest candidate (Conservative Robert Youngson) to 12,133,626 GBP for the richest candidate (Conservative Jacob Astor). The median wealth at death is 257,948 GBP. As a benchmark, the median gross value of the estate for males aged 65 and above in 2002 was 113,477 GBP,[20] indicating that the median candidate died with

These cases were left missing.

[19] We also discarded the very few "losing" candidates who eventually won a seat after 1970. Including them as winners or losers does not change the results (available upon request).

[20] All figures converted to real 2007 prices. Median wealth is computed from HM Revenue & Customs (HMRC) statistics table 13.2 "Estimated wealth of individuals in the U.K., 2002 (year of

almost twice the wealth of the median senior citizen in recent years. This result is roughly consistent with Gagliarducci et al. (2008) who find that the pre-parliament income of Italian politicians exceeds the median income in the rest of the Italian population by about 45 percent.

Given the well-known differences in social class between politicians from the two parties, it should not be surprising that Conservative candidates died significantly richer than their Labour counterparts. As shown in Table 4.1, the median wealth among Conservatives exceeded that among Labourites by $50,000$ GBP. Table 4.1 also provides the first indication that Conservative MPs died much wealthier than unsuccessful Conservative candidates; the median Conservative MP died with $483,448$ GBP while his unsuccessful counterpart passed away with a "mere" $250,699$ GBP. The difference on the Labour side is less than $10,200$ GBP. Figure 4.2 provides another look at this comparison by depicting the estimated density of log wealth for successful and unsuccessful candidates from each party. The first three wealth distributions (for winning and losing Labour candidates and losing Conservatives) look quite similar, but the wealth distribution for Conservative MPs appears to be shifted quite markedly upwards. Clearly, this difference must reflect either a substantial effect of office on wealth for Conservatives or a strong electoral bias toward wealthier candidates among Conservatives (or both).

4.4 Estimating the effect of office on wealth

Since political office is not randomly assigned among candidates, MPs and losing candidates may differ in ways that are correlated with both wealth and the probability of gaining office.[21] As noted in the previous section, our first line of defense against

death basis)," which uses the estate multiplier method to estimate wealth from probate values.

[21]The most obvious reason why winners and losers might systematically differ is that voters choose winners in a democracy, and voters might have preferences over candidate characteristics that are correlated with wealth. A more subtle, but probably more powerful, reason is that higher-quality candidates are likely to run in more favorable districts. Because the opportunity cost of running for office is presumably higher for wealthier and abler individuals, higher-quality candidates are likely

these confounding factors is to restrict our sample to relatively competitive candidates. In this section we describe statistical approaches we use to address remaining confounders.

4.4.1 Matching Estimates

Our dataset includes an unusually rich set of covariates for each candidate, which makes it possible to condition on many potential differences between winners and losers. In particular, for every candidate we record the year of birth, gender, party, schooling, university education, detailed occupation, titles of nobility, [22] and year of death. Descriptive statistics for the covariates are presented in Table 4.2. All characteristics except the year of death and wealth are measured from the *Times Guide to the House of Commons* biography that appears for the first constituency race of each candidate. The covariates are therefore "pre-treatment" in the sense that they are not affected by whether the candidate won office.[23]

To clarify the assumptions for the estimation let W_i be a binary treatment indicator coded one if candidate i served at least one period in the House of Commons, and zero if candidate i never attained office. X is an $(n \times k)$ matrix that includes our

to run in districts where the probability of winning is higher. If that is the case, winning candidates might die richer than losing ones even if voters ignore candidate characteristics *and* office has no effect on wealth. This more subtle selection effect may have been present in Britain in the period we examine because, with no residency requirement for being staged in a particular constituency, would-be candidates sometimes auditioned in multiple constituencies in a quest for the safest districts (Rush 1969). However, given our focus on close races this is presumably much less of a concern. In fact, we show below that in our sample there is no strong correlation between the vote share margin and wealth at death.

[22] We indicate that the candidate has a title of nobility if "Sir", "Viscount", "Lady" or "Lord" precedes the name in the *Times* biography.

[23] One question is whether we should condition on the year of death or not given that it is measured post-treatment and may be affected by wealth and political office. Below we report estimates including the year of death, but excluding it does not change the results (available upon request). The direction of the bias introduced by including or excluding year of death as a covariate is somewhat ambiguous. Candidates who lived longer may have had more time to make money, but on the other hand they may have drawn down their savings further; winning office, on the other hand, may lead to longer life or it may bring stress and an earlier demise. In separate tests, we find no systematic effect of gaining office on longevity, which suggests that post-treatment bias is not a concern.

k observed covariates for all n candidates with row X_i referring to the characteristics of candidate i. The variables $Y_i(0)$ and $Y_i(1)$ represent the wealth that candidate i would realize with and without gaining political office (i.e., "potential outcomes"). Evidently, only one of the potential outcomes is observed for each candidate. In the following we proceed by assuming unconfoundedness given the observe covariates, i.e. $(Y_1, Y_0) \perp W|X$, and common support so $0 < Pr(W = 1|X) < 1$ holds with probability one for (almost) every value of X (Rosenbaum & Rubin 1983).

The validity of the unconfoundedness assumption depends on the quality of the covariates in capturing the assignment mechanism. Arguably our unusually rich set of covariates captures the most obvious confounders. To the extent that wealthier candidates were better able to attain office (perhaps by using their connections to be placed in more favorable districts), the omission of wealth at the time of candidacy may be particularly problematic. However, while we do not measure pre-existing wealth explicitly (no such data is available), many of our covariates – such as whether a candidate was schooled at Eton, studied at Oxbridge, worked as a barrister, or has a title of nobility – will be highly correlated with pre-existing wealth and therefore indirectly control for this omitted factor. Later in the paper, we employ a different estimation strategy based on a regression discontinuity design that relies on close elections to control for unobservable factors.

We chose matching as our main method of covariate adjustment in order to avoid parametric assumptions and to keep the analysis transparent.[24] Specifically, we employ Genetic Matching (with replacement) following Diamond & Sekhon (2006)[25] with

[24] See Imbens (2004) and Rubin (2006) for reviews. We have tried several other techniques for covariate adjustment such as propensity score matching or regular Mahalanobis distance matching, weighting on the propensity score, and subclassification. All of these techniques lead to very similar results (available upon request).

[25] For each candidate we pick the M nearest neighbors according to the following distance metric

$$d(X_i, X_j) = \{(X_i - X_j)'(S^{-1/2})'WS^{-1/2}(X_i - X_j)\}^{1/2}$$

where W is a $(k \times k)$ positive definite weight matrix with zero in all elements except the main diagonal and $S^{1/2}$ is the Cholesky decomposition of S, the variance-covariance matrix of X. Notice that the only difference between this approach and regular Mahalanobis distance matching is the use of a generalized weight matrix W. If each of the k parameters in the diagonal of W are set equal to 1, $d()$ is the Mahalanobis distance. In Genetic Matching, the weights in the diagonal of W are chosen post-matching regression adjustment as proposed in Abadie & Imbens (2007). For comparison we also provide results from a regular OLS regression in the un-matched data. Since the above findings suggest that the effect of political office on wealth may depend on party, we conduct all estimations separately for each party.

Matching Results for the Conservative Party

The upper panel of Figure 4.3 presents the balance results for the Conservative party using one-to-one matching (i.e. $M = 1$). For each covariate, we plot the standardized bias as measured by the difference in means between the two groups scaled by the pooled standard deviation. Accordingly, circles to the right (left) of the dashed vertical line at zero indicate a higher incidence of a certain characteristic in the group of winning (losing) candidates. As expected, there are clear differences in the distribution of pre-existing characteristics between the two groups before matching (unfilled circles). MPs were more likely than unsuccessful candidates to have aristocratic backgrounds and elite educations. Winning candidates were also less likely to be in white-collar professions (engineering, accounting, or public relations), journalism, and teaching professions, and also less likely to have business backgrounds. After matching, however, we achieve a very high degree of covariate balance (filled circles). The standardized bias is now within 0.1 for all variables. The lowest p-value across paired t-tests and KS tests is 0.16, which indicates that the corresponding distributions for the matched groups are similar across all covariates. The two matched groups have very similar observed characteristics, such that any remaining difference between the wealth of winning and losing candidates can plausibly be attributed to the effect of treatment rather than pre-existing differences.[26]

The upper panel in Table 4.3 displays our effect estimates. The first column

by an evolutionary algorithm such that balance across treatment and control groups is maximized. Balance is measured by the lowest p-value across covariate-by-covariate paired t-tests for differences in means and bootstrapped Kolmogorov-Smirnov tests for the equality of distributions. See Sekhon (2007) for details.

[26]Notice that there are no union officials or miners among the Conservative candidates so these two variables are balanced in the unmatched data already.

presents the results from a simple OLS regression (with robust standard errors) of wealth on the treatment indicator including all the covariates. Columns two and three display the results from the matching estimator for two quantities of interest: The average treatment effect (ATE) given by $\tau_{ATE} = E[(Y_i(1) - Y_i(0)]$ and the average treatment effect for the treated (ATT) given by $\tau_{ATT} = E[(Y_i(1) - Y_i(0)|W_i = 1]$ with Abadie & Imbens (2006) standard errors. Across specifications, we find a robust and substantial impact of serving on wealth at the time of death. We estimate that serving in Parliament increased wealth at death by between 71 and 155 percent, depending on the specification. For all specifications we soundly reject the null hypothesis of no effect at conventional levels.

Matching Results for Labour Party

Balance results for Labour candidates are reported in the lower panel of Figure 4.3. Again, we find some pronounced differences in the covariate distributions between MPs and unsuccessful candidates before matching. The discrepancies between winners and losers are roughly the reverse of those for the Conservative party: among the winning Labourites there is a smaller fraction of candidates with an Oxbridge education, Eton schooling, or business background than among the unsuccessful candidates, but a higher fraction of union officials and local politicians. After matching, these differences are almost completely removed. We obtain a very high degree of balance on all covariates, with the lowest p-value across all balance tests being .30.

The lower panel in Table 4.3 presents the matching-based effect estimates for Labour candidates. Consistent with the distributional box-plots shown earlier, we find no effect of serving on wealth at death. The point estimates across all models are close to zero. Although this null finding is not very precisely estimated, the difference between the effect for Conservative and Labour MPs is clear: in an OLS regression pooling the two parties, the p-value on the test that the coefficient is the same for the two parties is 0.05.

4.4.2 Regression Discontinuity Design Results

The matching results presented so far rest on the assumption of unconfoundedness, which fails if, conditional on the observed covariates, there remain imbalances in important unobserved factors between winners and losers. Controlling for unobserved confounding is impossible in most observational studies, but the unique nature of political contests provides an opportunity to apply a regression discontinuity (RD) design to the problem (Thistlethwaite & Campbell 1960). Following pioneering work by Lee (2008), we note that in very close elections, the assignment to political office is largely based on random factors. While winning candidates may generally be different from losing candidates at the time of the election (e.g., better looks, more money, or greater speaking ability), there is no reason to expect the winners and losers of elections decided by razor-thin margins to systematically differ in any way. The RD design therefore attempts to estimate the difference in wealth precisely at the threshold where winners and losers are decided, i.e. where the margin of victory approaches zero. If local random assignment holds at the threshold, the RD estimate can thus be as credible as an estimate from a randomized experiment.

In particular, let Z_i be the vote margin for candidate i. For winning candidates, Z_i is computed from their first successful race as the difference between their own vote share and that of the runner-up. For losing candidates, Z_i is computed from their best race as the difference between their vote share and that of the winner.[27]

Given this definition, gaining office is a deterministic function of the margin $W_i =$

[27]The application of a regression discontinuity design to a candidate-level outcome such as wealth requires addressing the fact that many candidates stand for election more than once, and thus losers sometimes reappear as winners in later elections. Our approach obviates the resulting compliance problems (Angrist, Imbens & Rubin 1996) by defining the assignment variable in the context of a candidate's entire electoral history: the best race for losers and the first successful race for winners. This definition implies that close winners will be compared to the most competitive losers available. As our balance tests later show, close winners and losers defined in this way do not differ in any observed covariate, including the number of previous races the candidate has run. We have conducted additional tests using a fuzzy regression discontinuity design, which uses success in a candidate's first race as an instrument for serving in Parliament. The point estimates are similar but very imprecise given our limited sample size and the efficiency loss incurred. The fuzzy design is particularly inefficient in the setting of UK elections because new candidates are often staged in unwinnable districts in order to gain experience, which means that the first race provides only a very noisy signal of candidate quality.

$1\{Z_i \geq 0\}$.[28] In other words, all candidates with $Z_i > 0$ are assigned to the group of winners and enter Parliament while candidates who score just below the threshold are assigned to the group of losing candidates and do not enter Parliament. The average treatment effect at the threshold $Z = 0$ is then defined as

$$\tau_{RDD} = lim_{z \downarrow 0} E[Y_i | Z_i = z] - lim_{z \uparrow 0} E[Y_i | Z_i = z] = E[Y_i(1) - Y_i(0) | Z_i = 0] \quad (4.1)$$

which is identified under the assumption that $E[Y(0)|Z = z]$ and $E[Y(1)|Z = z]$ are continuous in z.[29] This assumption is fairly weak and will fail only if candidates can strategically sort around the threshold. In fact, Lee (2008) shows that as long as the vote share includes some random component with a continuous density, treatment status is randomized at the threshold of winning.[30]

The upper panel in Figure 4.4 presents the graphical results from the RD design for Conservative candidates. Wealth is plotted against the vote share margin (Z_i). The dotted vertical line at zero indicates the threshold separating MPs (to the right of the threshold) and unsuccessful candidates (to the left of the threshold). The solid lines represent the conditional expectation functions of wealth given the vote share margin approximated using a locally weighted polynomial regression fitted to both sides of the threshold; pointwise .95 confidence bounds are indicated by dashed lines. Recall that the effect of office on wealth in the RD design is defined as the difference of the two conditional expectation functions at the threshold. By (minimally) extrapolating the polynomial fit to the threshold, we estimate that marginal winning candidates died

[28] There are no ties in our data.

[29] Notice that compared to the matching estimates shown above, unconfoundedness holds trivially here since W does not vary conditional on Z, but the overlap assumption is violated because the probability of assignment is either $Pr(W_i = 1|Z_i > 0) = 1$ or $Pr(W_i = 1|Z_i < 0) = 0$ depending on whether a candidate scores below or above the threshold.

[30] As is well known, the RD design is likely to have a very high degree of internal validity, but we pay a price in terms of decreased external validity and also efficiency. τ_{RDD} is a local average treatment effect informative only for marginal candidates close to the threshold of winning (unless additional homogeneity assumptions are introduced). This is desirable in our context, however, because the counterfactual seems more reasonable for marginal compared to "unbeatable" candidates. Moreover, given that candidates in closer races attract more public scrutiny and face a higher risk of electoral defeat, rent seeking may be limited compared to candidates in safe districts (Barro 1973, Besley & Case 1995, Besley & Burgess 2002). Presumably our estimates of the returns to office therefore provide a conservative lower bound for the average across all MPs.

with about 546,000 GBP compared to about 298,000 GBP for losing candidates. The first column in table 4.4 displays the formal estimate of this jump in the conditional expectation function at the discontinuity which is about 250,000 GBP or about a 83 percent increase in wealth at death. The (non-parametric) bootstrapped .95 percent confidence interval ranges from 8 to 212 percent. This estimate is similar to the matching results obtained earlier and suggests that narrowly successful Conservative candidates almost doubled their wealth by winning office.

Another notable feature in the upper panel of Figure 4.4 that the conditional expectation of wealth is not steeply increasing in the vote share margin over the support of the vote share variable. Assuming that post-treatment wealth is highly correlated with pre-existing wealth (i.e. wealth at the time a candidate ran for office) this would provide evidence against the claim that candidates could simply buy office via placement in very safe seats or were otherwise strongly selected based on existing wealth (at least for our sample of competitive winners and losers). This might explain why the RD results do not differ much from the regression and matching findings presented earlier.

The lower panel in Figure 4.4 displays similar graphical results for the Labour candidates. Again, the RD findings correspond very closely with the matching results. There is almost no discontinuity at the threshold, suggesting that there is no effect of winning office on wealth among Labourites. The third column in table 4.4 displays the estimate of the jump in the conditional expectation function at the discontinuity which is about 56,000 GBP or about a 18 percent decrease in wealth at death. The bootstrapped .95 percent confidence interval ranges from -52 to 32 percent.

As expected, the results from the graphical analysis do not change when we introduce covariates into the estimation. To formally estimate the difference of the two regression functions at the discontinuity point while including our full set of covariates, we follow the proposal by Imbens & Lemieux (2007) and fit a local linear regression of the form:[31]

$$\min_{\alpha,\beta,\tau,\gamma,\delta} \sum_{i=1}^{N} 1\{-h \leq Z_i \leq h\} \cdot (Y_i - \alpha - \beta \cdot Z_i - \tau \cdot W_i - \gamma \cdot Z_i \cdot W_i - \delta' X_i)^2 \quad (4.2)$$

where τ identifies our treatment effect estimate. The variance of τ can simply be estimated using the standard robust variance from the OLS regression. The bandwidth around the threshold of winning, h, is chosen by the Imbens and Lemieux (two-sided) cross-validation criterion.[32] The optimal bandwidth according to this criterion is about 15 percentage points of vote share.[33] The second and fourth columns in Table 4.4 present results for this regression with our full set of covariates (including schooling, university education, occupation, gender, year of birth, and year of death). Just as in a randomized experiment the inclusion of covariates has only a small effect on the estimate of τ because, in the close neighborhood of the threshold, all observed and unobserved covariates should be independent of W. We again reject the null at the conventional levels but the standard errors, as expected, are slightly larger than in the matching analysis because the RD approach focuses on the neighborhood of the threshold, where there are fewer observations.

[31]See Imbens & Lemieux (2007) for a discussion of alternative estimation strategies. They key issue is that the RD estimand is a single boundary point, so that nonparametric kernel regression may contain a high order bias due to slow convergence. Local linear regression provides a practical solution to this problem.

[32]Imbens & Lemieux (2007, equation 5.12).

[33]As suggested by the flatness of the conditional expectation, our results are fairly insensitive to the choice of bandwidth for the rectangular kernel, although obviously the standard errors tend to increase as the bandwidth is decreased due to the smaller number of observations. For example, for the Conservatives the estimated treatment effect (including all covariates) is .82 (.59) when we use half the optimal threshold (i.e. 7.5 percentage points) and .57 (.29) when double the optimal bandwidth (i.e. 30 percentage points) is used. For completeness, the same estimates without all covariates are .71 (.45) for half and .63 (.27) for double the bandwidth.

4.4.3 Robustness Tests for RD Estimation

Test for Wealth Jumps at Non-discontinuity Points

Following the proposal by Imbens and Lemieux (2007), we test for jumps in wealth at points other than the threshold at which office was assigned. We produce RD estimates at 5 percentage point increments along the range of the vote share variable, in each case limiting analysis to either the winning or losing candidates.[34] Figure 4.5 compares these placebo effect estimates with our estimate of the effect of winning office on wealth. (We focus on Conservative candidates, since we did not find an effect for Labour.) The upper panel presents the point estimates for each of the placebo runs contrasted with the estimate at the true threshold; the lower panel presents the corresponding t-values. The true effect estimate clearly stands out from the placebo effects. The placebo effects are generally smaller in magnitude; all of the them are highly insignificant at conventional levels. This finding increases our confidence that our estimate measures the effect of gaining office rather than a random artifact of the data.

Test for Zero Average Effect on Placebo Outcomes

Here we assess whether winning office appears to have affected candidate characteristics (such as year of birth) that could not possibly have been affected by serving in Parliament. This type of test, which was first applied in an RD setting by Lee, Moretti & Butler (2004),[35] looks for evidence that the winners of very close elections do not appear to have been randomly selected; if they were, we would expect to see no treatment effect on these placebo outcomes. We repeatedly obtain RD estimates at the threshold between losers and winners, where instead of wealth as the outcome we used each of our covariates in turn. Table 4.5 displays the results for both par-

[34]By focusing on each subsample separately, we follow Imbens & Lemieux (2007, pg. 27), who note that otherwise our regression function would assume continuity at a point where we know there is a break.

[35]See Imbens & Lemieux (2007) for a discussion.

ties. The 95% confidence interval on the estimated placebo effect includes zero for all covariates over both parties, with only one exception (an indicator for candidates whose secondary school is not reported in their bios). After correcting for multiple comparisons no differences are significant at the threshold; the degree of imbalance across groups is similar to what we would expect in a randomized experiment.

Included in Table 4.5 with the covariates we considered previously are two additional measures that we judged to provide a further useful indication of whether candidates might somehow be sorting around the threshold. One such measure is the vote share for the candidate's party in the same district in the prior election (indicated by "Previous VS"). Since candidates competed to be staged in favorable districts, this is likely to be a good measure of the desirability of the seat and therefore the quality of the candidate. The second measure is the number of attempts the candidate took before the decisive race (i.e. the first winning race for winners or the best losing race for losers), indicated by "Previous Attempts" in Table 4.5. If the winners in our dataset triumphed through persistence, we would expect this covariate to systematically differ between the two groups. The fact that we do not find a significant difference for either variable provides support for the validity of the identification strategy.

4.5 Discussion

Based on the analysis in the previous section, we conclude that serving in the House of Commons roughly doubled the wealth at death of Conservative candidates on average but had no effect for candidates of the Labour party. It remains to consider possible channels by which serving Parliament could have such a strong, party-specific effect on personal wealth.

4.5.1 Did MPs Make their Money in Office or After Retiring?

As a starting point, we examine our data for evidence of when Conservative MPs made their money – while sitting in Parliament or after retiring (see Table 4.7). We first regressed log wealth on the total time the MP lived after being elected (denoted "Years as MP and Former MP"), as well as the MP's year of birth, margin of winning (in first successful race), and indicators for whether the MP attained front bench or cabinet positions and attended elite educational institutions (reported in column 1). The regression indicates that MPs who had longer careers as MPs and ex-MPs (i.e. who lived longer after being elected) died with more money (p-value = .03). The point estimate suggests that living an additional year after entering office (or, equivalently, entering office one year earlier) is correlated with about a 2.9 percent increase in wealth. In a similar vein, the dummy variable for front bench or cabinet service enters positively and with a substantial magnitude, although we lack sufficient precision to reject the null at conventional levels. The coefficients on the other control variables have the expected positive signs. Consistent with Figure 4.4, a candidate's margin of winning and wealth at death are not significantly correlated, which speaks against the idea that wealthier candidates secured spots in more favorable districts or were otherwise favored by the electoral process.

In an attempt to disentangle money made in office and after retiring, we ran a further regression (reported in column 2) in which we separated post-election years into "years served as MP" and "years lived as a former MP." We find that an additional year in office is associated with a roughly 3% increase in wealth at death (p-value = .02). This suggests that longer service, which presumbaly comes with more valuable political knowledge and connections, translates into higher wealth. We also find that an additional year after retirement from office is associated with a roughly 2% increase in wealth at death but this estimate is not significant (p-value = .19). This is consistent with the interpretation that a considerable share of the financial benefit of office came while an MP was sitting in Parliament and not only after his or her retirement. It also could be consistent with the interpretation that MPs made their money after leaving Parliament, and that their post-office earnings depended on the extent of their parliamentary experience. Serving in Parliament could thus be thought of as an investment in human capital, like education, whose payoff (collected after "graduating" from government) is increasing in years served. (In the U.S. Diermeier et al. (2005) find that congressional experience significantly raises post-congressional wages both in the private and the public sector.) To test this idea we conduct a third regression where we interact years in office and years out of office. If serving in Parliament was indeed an investment in human capital and boosted wealth primarily by making post-parliamentary employment more valuable, we would expect a substantial positive coefficient on the interaction term and a much smaller coefficient on "years as MP." In fact, as indicated by column 3 of Table 4.7, the interaction term is essentially zero and highly insignificant while the magnitude of the "years as MP" coefficient does not diminish. We calculate that an additional year lived as a former MP is associated with a 1.8 percent (p-value= 0.21) increase in wealth at death for a person who served 10 years as an MP, compared to a 2.2 percent (p-value=.41) increase for somebody who served 30 years as MP. This suggests that extra years in Parliament did not increase wealth primarily by raising post-office earnings, but rather that an MP's years in office were themselves lucrative.

4.5.2 How Did MPs Make Money In Office?

One possibility to address immediately is that MPs' official pay explains the financial benefit of office: perhaps Conservative MPs received a significantly higher salary than what they would have earned outside of Parliament. This conjecture is completely at odds with the evidence, however. Not only was the MP salary modest compared to wages in professions MPs commonly pursued before entering office,[36]

[36]Data from a survey conducted among new members of Parliament in 1979 indicate that over three-quarters of entering MPs took a pay cut to serve in Parliament; at a time when an MP's salary was 6,897 GBP, the median backbencher had left a job paying 11,000 GBP (Judge 1984, pg, 68). The New Earnings Survey, which was first conducted in 1971, indicates that over the last several decades MPs have consistently earned somewhat more than journalists and university professors but

Conservatives were more likely to face a pay cut after being elected, given that they tended to come from lucrative careers in law and business. If salaries were the dominant factor, we might expect to see the union officials, journalists, and lecturers of the Labour Party profit, but not the accountants, barristers, and managing directors of the Conservative party. Given that we see the opposite, salary evidently does not explain the observed pattern of benefits from office.

It is also unlikely that health effects can explain our findings. If the status boost of serving in Parliament improved health (see Redelmeier & Singh (2001), but also Sylvestre, Huszti & Hanley (2006)), it may have extended MPs' working lives and increased the size of their estates. (On the other hand, living longer can deplete savings.) In fact we find no difference in the longevity of MPs and unsuccessful candidates. For both parties, a treatment indicator for winning office is statistically insignificant in regressions of either age at death or year of death (including all our covariates). Moreover, in our balance tests for the regression discontinuity design, we found that there is no discontinuity in year of death at the threshold of winning (see Table 4.5). Finally, none of our results are affected by including the year of death in the regressions.

The most obvious channel through which winning office may have increased wealth is through lucrative, politically linked outside employment (particularly directorships and "parliamentary consultancies") that have periodically come under public scrutiny. We therefore ask whether MPs sat on more boards of directors than unsuccessful candidates with similar backgrounds. We used the *Directory of Directors*, an annual listing of the directors serving on boards of companies traded on the London Stock Exchange, to count the number of directorships listed in 1983 for each of the candidates for whom we also collected wealth data. We find that, controlling for our standard battery of covariates (gender, year of birth, year of death, and indicators for schooling, university and titles of nobility), Conservative MPs indeed had significantly ($p = .08$) more directorships than unsuccessful candidates, with the predicted number of directorships being .46 for losers and .13 for losers at covariate means. Among

less than legal professionals and managers in large companies.

Labour MPs, we find that losers actually had more directorships than winners (.54 vs .10), although the difference was not significant (p = .47) and was driven largely by a single outlier among the losers who held 19 directorships. (With that outlier removed the expected rate is .09 for losers and .11 for winners (p = .80).)

It seems worth asking how much of the total wealth gain we estimate for Conservative MPs could be accounted for solely by the politically linked directorships that MPs collected. To answer this, we conduct the following back-of-the-envelope calculations, which suggest that directorships could in fact have accounted for a considerable portion of the wealth benefit of serving in Parliament.

First we need to determine by how much the added directorships would be expected to increase the average earnings of MPs compared to unsuccessful candidates. As noted above, winning office was expected to increase the average number of public-company directorships among Conservatives from .13 to .46. These figures considerably understate the increase in all directorships, however, because they include only directorships of public companies, whereas many MPs held private-company directorships. If we scale the increase in 1983 directorships according to the total number of directorships reported in the 1975 RMI, it works out to an increase in roughly one directorship per member. The average annual fee for outside directors was about 25,000 GBP (in 2007 prices) plus benefits (Hollingsworth 1991, pp. 21, 157), indicating that winning office conferred roughly that amount in extra directorship income on our sample of MPs, at least in the 1970's and 1980's.

Next we need to convert the gain in wealth at death that we estimate above into a difference in annual earnings. As noted previously, we estimate the average wealth benefit of serving in Parliament for our sample at about 250,000 in 2007 GBP. Only a fraction of earnings ultimately is bequeathed; using U.S. probates from the 1960's and 1970's Menchik & David (1983) estimate the marginal propensity to bequeath from earnings at about .25 for the top quintile of his sample. If this data is an appropriate rough guide in our context, MPs would have had to earn roughly 1 million pounds more (at 2007 prices) on average over the course of their lifetimes compared to unsuccessful candidates in order to boost their estates by the estimated amount.

Since the median Conservative MP served 18 years and lived 17 more, this would require earning around 25,000 GBP more per year after being elected than one would have earned outside of politics. This is precisely the boost in annual directorship fees in 1983 that we estimated above. Recognizing that this calculation is necessarily quite rough, it does seem that directorships alone can account for a sizable proportion of Conservative MPs' wealth gains from being elected to Parliament.

Evidence from several countries suggests that employing sitting or former politicians as board directors and or consultants is valuable to firms precisely because of the political connections and knowledge that (former) legislators possess ((Ferguson & Voth 2008, Goldman et al. 2008*a*, Goldman et al. 2008*b*, Jayachandran 2006, Faccio 2006, Khwaja & Mian 2005, Johnson & Mitton 2003, Fisman 2001, Roberts 1990). Similarly, in the British case plenty of anecdotal evidence suggests that what MPs provided for their clients as directors and parliamentary consultants is a form of influence more subtle than votes, exerted via personal connections to ministers and members of the civil service, as well as information about the affairs of government. MPs and the outside interests who retain them have at times been quite candid about the nature of this political exchange. A month after leaving office as Chancellor of the Exchequer in the wake of the 1964 general election (and while a sitting MP), Reginald Maudling accepted a position as executive director of a merchant banking firm, with fees estimated at over five times his MP salary. Journalist Andrew Roth noted that "the firm made it clear to the financial writers present that it was very useful indeed to have on tap the knowledge and contacts made by a former Cabinet Minister who had been Chancellor of the Exchequer and President of the Board of Trade" (Roth 1965, pg. xii). In 1968, Conservative MP Anthony Courtney explained that "Election to the House of Commons not only consolidated but also improved my business affairs. I had acquired for the benefit of the firms with which I was connected improved personal contact with the Board of Trade and other ministers" (Courtney 1968, pg. 63). Muller (1977) concludes that Labour MPs acted as "servants," "spokesmen," and "consultants" for the unions that supported them. Conservative MP Sir Peter Emery, who served 42 years in office, was twice censured by the

House of Commons Public Accounts Committee in 1980 for gross profiteering by his company on Government contracts. He was also criticized for using a visit to Ghana as a member of a parliamentary delegation to negotiate privately a road-building contract for a British firm with which he had business links. Emery accumulated several directorships during his political career including one with Winglaw Property Group, for which the newspaper *The Observer* reported 500,000 GBP in bonuses for property deals in 1989-92 alone.[37] After resigning from his position as Defence and Environment minister in 1981, Sir Markus Fox, who served as a Conservative MP form 1970 to 1997, and a fellow MP Keith Speed created the public affairs consultancy Westminster Communications Ltd., which according to the RMI from various years served a long list of clients with political lobbying.[38] Before the Select Committee on Members Interest, Fox stated "We thought if we, as Members of Parliament, were actually controlling the company we could ensure we only acted for those clients who we were convinced were of good standing." He also claimed that "There is a need for a lobbying industry. That is proven by the success of the companies, their growth and the fact that people are prepared to pay for this sort of information and the amount of time that is spent in often putting a good case forward."[39]

In sum, the evidence suggests that being elected to Parliament endowed politicians with valuable political connections and knowledge that (through directorships and other employment) helped special interests to influence policy and anticipate changes in policy. Given this, one may expect this political exchange to be more valuable for MPs with better access to political power. There is unfortunately no exact way to disentangle the degree to which Conservative MPs enjoyed access to power given that the Conservative party ruled for most of our time period. The seven elections we observe saw power change hands only three times: Conservatives won power in 1951, Labour won it back in 1964, and the Conservatives again triumphed in 1970.

[37]Obituary Sir Peter Emery, The Guardian, December 11, 2004. Obituary Sir Peter Emery, Telegraph, December 13, 2004. Also see Hollingsworth (1991, pg. 38-43,45-46).

[38]RMI of January 13th, 1986, RMI of December 14th, 1991, RMI of January 31st, 1997.

[39]Evidence to Select Committee on Members' Interest, HC 44-vii, p. 200-202. Also see Hollingsworth (1991, pg. 21,69,70-74,105-107,120-122).

As a rough test, we assessed whether members who entered Parliament in years when their party was in power enjoyed larger wealth benefits (compared to unsuccessful candidates whose best race was in the same year) than those who entered office when their party was not in power. Since we do not have enough data to sustain subgroup analysis with an RD design, we rely on a regression in which the effect of serving in office is allowed to vary based on whether the candidate ran in a year when the Conservatives attained office or not. We find that for Conservatives the effect of serving in office on wealth is much larger for MPs that entered in the years in with the Conservatives was in power. This result is consistent with the idea that better to access to political power strengthened the ability of MPs to translate their political knowledge and connections into higher wealth. We find no significant difference for the Labour candidates.

4.5.3 Why Did the Benefits of Office Differ by Party?

The question remains why Labour MPs did not appear to derive as large a financial benefit from office as did their Conservative counterparts. We argue that Conservative MPs and Labour MPs operated in essentially separate markets for political influence; the difference in the structure of those markets helps to explain why MPs in one market retained a greater amount of surplus than in the other.

The Labour and Conservative parties in the period we examine were organized and financed quite differently from each other, in ways that ultimately affected how MPs for each party related to outside interests. In the Labour Party, a small number of very large unions provided the bulk of the financing and exercised a corresponding amount of direct influence over policy and political representation. Between 1945 and the 1990s, unions consistently provided 80-90% of the funding of the Labour Party central office and around two-thirds of the party's funding overall (including local organizations) (Harrison 1960, Pinto-Duschinsky 1981, Pinto-Duschinsky 1990). Trade unions also directly provided a plurality of delegates to national party conferences as well as to local constituency councils responsible for selecting parliamentary candidates. By contrast, the Conservative Party drew its funding from a larger number of smaller players and political influence was correspondingly diffuse. Company contributions provided as much as 30% of the party's income overall, but those contributions came from several hundred different companies with fairly weak coordination among themselves.[40] The bulk of Conservative Party finance came from individual contributions, whether through party fundraisers held by local constituency organizations (which alone brought in more money than did corporate contributions) or large and undisclosed individual contributions and bequests (Pinto-Duschinsky 1981, Pinto-Duschinsky 1990, Fisher 1994).

Because unions were intimately involved in the selection of Labour candidates and in many cases financed their election to Parliament, Labour MPs tended to enter office with well-defined obligations to specific unions. The means by which unions ensured the loyalty of MPs was clearest in the case of direct sponsorships, an arrangement that was formalized in the party's 1933 "Hastings agreement." Between 1945 and 1975 sponsorships extended to over 30% of all Labour candidates and over 40% of all Labour MPs (Muller 1977, Harrison 1960). Unions sponsored parliamentary prospects as early as the candidate-selection stage; if a union's sponsored member were selected to stand for election (a process in which the unions jointly played a large role), that union would provide campaign finance through the election (Rush 1969). Unions tended to sponsor and promote candidates from their own ranks who were likely to remain loyal representatives once in office and return to the union bureaucracy after retirement from Parliament (Muller 1977). Occasionally, a sponsored member deviated from the position advocated by the sponsoring union, with the consequence that the MP lost the sponsorship and, often, subsequently the seat (Muller (1977, pg. 153), Harrison (1960)). By contrast, the process of selecting Conservative candidates was shared between the party's national office and local constituency committees, neither of which gave a particularly privileged role to individual companies or other outside groups (Rush 1969). Conservative candidates thus generally entered office with loyal-

[40] As a comparison of the distribution of union and corporate donations, in 1987 the political expenditures of the largest union (Transport and General Workers) to the Labour Party exceeded the combined political donations of 1,300 of Britain's largest companies to the Conservative Party Pinto-Duschinsky (1989, pg. 208).

ties to the party and local constituency committees but with no exclusive obligations to any particular outside interest.

We suggest that it was largely because the unions were effective in controlling politicians through non-monetary means that Labour MPs captured a relatively small economic bonus from serving in Parliament. Conservative MPs operated in an open market for political services. Because client firms were numerous and poorly organized among themselves, they competed for MP loyalty and paid substantial sums to secure it, largely through consulting and lobbying contracts and directorship positions. On the Labour side, the labor unions suppressed the market for MPs' services by controlling the party and, through the party, the politicians themselves. The trade unions' solution looks something like backward vertical integration: instead of purchasing political services on the open market, the unions created a subsidiary (the Labour Party and its MPs) to supply political goods. (In fact, the early history of the Labour Party is basically consistent with this interpretation (Beer 1965).)

In sum, we surmise that it is largely because business interests were less organized than the unions, and had less power in Conservative politics than did unions in Labour politics, that Conservative MPs profited more from office than did Labour MPs. It is likely that the value of office in a variety of contexts similarly depends on the extent to which constituents can use formal means of political control and are organized enough to restrain competition for political influence.

4.5.4 Did MPs' Outside Arrangements Affect Their Behavior?

Many analyses of the role of special interests in US politics examine links between campaign contributions made by interest groups and votes taken by members of Congress (Peltzman 1984, Kroszner & Strahan 1999). A similar investigation is not likely to be fruitful in Britain and many other political systems, where party control of MPs is higher and legislative votes almost always follow party lines.

While positions taken in parliamentary votes are not very informative about the influence of MPs' clients on policymaking, the attendance of MPs for voting sessions does suggest that outside interests tend to distract members from legislative work. As a simple test, Figure 4.6 compares the attendance rates (the percent of eligible votes personally attended or told) for the 2005-2007 period for MPs with and without outside interests as declared in the latest RMI.[41] We find that for both parties, MPs with outside interests (directorships, consultancies, and work in journalism) attended fewer votes; attendance rates are around 4-6 percentage points lower and the differences are all significant at conventional levels. The exception are directorships for Labour MPs where we cannot reject the null of no difference. We also find no such difference between MPs who did and did not carry on regular outside employment (such as work as a barrister, medical doctor, etc.).

These findings are consistent with analysis by Muller (1977), who examines the degree of legislative participation by sponsored and non-sponsored Labour MPs. He finds that sponsored MPs were more active than other MPs on issues close to the interests of their sponsors (e.g. mining or railway issues), but that on the whole they were less active members of Parliament, participating in question time, standing committees, and debates far less than non-sponsored members. The results are also consistent with Gagliarducci et al. (2008) who find that in the Italian Parliament, politicians with more outside income are less committed to parliamentary activity measured by their voting attendance and the number of proposed bills.

4.6 Conclusion

Many studies have shown that private firms gain from connections to politicians, but little is known about how politicians benefit from firms and other groups seeking political connections (Merlo 2006, pg. 33). If there is indeed a "gift exchange" (Choi & Thum 2007, pg. 22) between politicians and politically connected firms, one

[41] Abstention rates are from `http://www.publicwhip.org.uk` (retrieved May 5, 2008). We ignore MPs who hold office as speaker and are thus not allowed to vote. Notice that several other factors may contribute to low attendance rates such as absence on constituency business, delegations to international organisations, illness, bereavement, or paternity/maternity leaves.

can expect politicians to benefit financially from office just as firms do from connections to officeholders. However, this perspective has been largely overlooked so far, presumably because estimating the financial benefits of political office is challenging empirically.

In this paper we measure the value of political power in post-war British politics using data about the estates of British politicians who entered the House of Commons between 1950 and 1970 and often served well into the 1990s. We identify the effect of office on wealth at death both by explicitly controlling for a wide variety of candidate-level characteristics and by employing a regression discontinuity design that exploits the quasi-random assignment to office that takes place in close district races. We find that serving in Parliament almost doubled the wealth of candidates of the Conservative Party, but had no appreciable effect for Labour candidates. These financial benefits of office are likely attributable to payments from private firms to sitting legislators and (perhaps less so) lucrative employment opportunities provided to politicians after retirement. Conservative MPs financially benefited from directorships and consulting work that accrued to them as a result of serving in political office, and the benefit they enjoyed appears to have been larger when their party was in power. Labour politicians had explicit relationships with unions that were far less lucrative; we surmise that Labour MPs were paid less for political services because the trade unions were better organized and secured their services largely through non-monetary means.

While our application benefits from data resources unique to the UK, our general approach is broadly applicable and could be used to measure the political power premium in other political systems. Faccio (2006) shows that the strength and scope of political connections, as well as the benefits of these connections to firms, vary widely across countries. One may expect the political power premium to vary based not only on these features but also on the organization and financing of political parties, the degree of legislator independence (both from party leadership and from specific interest groups), and the extent of restrictions on legislator conflict of interest.

4.7 Tables for Chapter 4

Table 4.1: Gross Wealth at Death (Real 2007 GBP) for Competitive Candidates Who Ran for the House of Commons Between 1950-1970 (Estimation Sample)

	Mean	Min.	1st Qu.	Median	3rd Qu.	Max.	Obs
Both Parties:							
All Candidates	599,385	4,597	186,311	257,948	487,857	12,133,626	427
Winning Candidates	828,379	12,111	236,118	315,089	722,944	12,133,626	165
Losing Candidates	455,172	4,597	179,200	249,808	329,103	8,338,986	262
Conservative Party:							
All Candidates	836,934	4,597	192,387	301,386	743,342	12,133,626	223
Winning Candidates	1,126,307	34,861	252,825	483,448	1,150,453	12,133,626	104
Losing Candidates	584,037	4,597	179,259	250,699	485,832	8,338,986	119
Labour Party:							
All Candidates	339,712	12,111	179,288	250,329	298,817	7,926,246	204
Winning Candidates	320,437	12,111	193,421	254,763	340,313	1,036,062	61
Losing Candidates	347,934	40,604	177,203	243,526	295,953	7,926,246	143

Table 4.2: Characteristics of Competitive Candidates Who Ran for the House of Commons Between 1950-1970 (Estimation Sample)

	Mean	SD	Mean	SD	Min	Max	
Teacher	0.11	0.32	Female	0.05	0.21	0	1
Barrister	0.10	0.30	Year of Birth	1919	9.68	1890	1945
Solicitor	0.07	0.25	Year of Death	1995	6.40	1984	2005
Doctor	0.02	0.15	Schooling: Eton	0.06	0.24	0	1
Civil Servant	0.01	0.11	Schooling: Public	0.30	0.46	0	1
Local Politician	0.25	0.43	Schooling: Regular	0.39	0.49	0	1
Business	0.14	0.35	Schooling: Not reported	0.25	0.43	0	1
White Collar	0.10	0.30	University: Oxbridge	0.28	0.45	0	1
Union Official	0.02	0.15	University: Degree	0.36	0.48	0	1
Journalist	0.10	0.30	University: Not reported	0.36	0.48	0	1
Miner	0.01	0.08	Title of nobility	0.03	0.17	0	1

Note: All covariates except year of death are measured at the time of the candidates' first race between 1950-1970. The minimum and maximum value for all the variables in the first column is zero and one respectively.

Table 4.3: Matching Estimates: The Effect of Serving in the House of Commons on (Log) Wealth at Death

	Conservative Party			Labour Party		
	OLS ATE	Matching ATE	Matching ATT	OLS ATE	Matching ATE	Matching ATT
Effect of Serving	0.54	0.86	0.95	0.16	0.14	0.13
Standard Error	0.20	0.26	0.34	0.12	0.18	0.15
Covariates	x	x	x	x	x	x
Percent Wealth Increase	71	136	155	17	15	13
95 % Lower Bound	15	41	31	-6	-19	-15
95 % Upper Bound	153	293	398	48	63	52

Notes: N=223 for the Conservative party, N=204 for the Labour party; for the ATT estimation there are 104 treated units for the Conservative party and 61 for Labour. Covariates include all covariates listed in table 4.2. ATT=Average Treatment Effect for the Treated, ATE=Average Treatment Effect, OLS=Ordinary Least Squares. Matching results are from 1 : 1 Genetic Matching with post-matching regression adjustment. Standard errors are robust for the OLS estimation and Abadie-Imbens for matching.

Table 4.4: Regression Discontinuity Design Results: The Effect of Serving in the House of Commons on (Log) Wealth at Death

	Conservative Party		Labour Party	
Effect of Serving	0.61	0.66	-0.20	-0.25
Standard Error	(0.27)	(0.37)	(0.26)	(.26)
Covariates		x		x
Percent Wealth Increase	83	94	-18	-23
95 % Lower Bound	8	-7	-52	-65
95 % Upper Bound	212	306	31	71

Note: Effect estimates at the threshold of winning $\tau_{RDD} = E[Y(1) - Y(0)|Z = 0]$. Estimates without covariates from local polynomial regression fit to both sides of the threshold with bootstrapped standard errors. Estimates with covariates from local linear regression with rectangular kernel (equation 4.2); bandwidth is 15 percentage point of vote share margin with robust standard errors. For the Conservative party N=223 for the estimates without covariates and N=165 with covariates. For the Labour party N=204 for the estimates without covariates and N=164 with covariates

Table 4.5: The Effect of Serving on Placebo Outcomes

	Conservative Party			Labour Party		
Placebo Outcome	Placebo Effect	95. UB	95 LB	Placebo Effect	95. UB	95 LB
Year of Birth	2.79	8.10	−2.62	2.50	8.62	−3.77
Year of Death	2.08	5.97	−1.89	2.23	6.23	−1.91
Female	−0.01	0.14	−0.16	−0.03	0.06	−0.12
Teacher	−0.09	0.06	−0.23	−0.23	0.01	−0.47
Barrister	0.09	0.25	−0.09	−0.07	0.05	−0.18
Solicitor	−0.13	0.07	−0.33	0.03	0.15	−0.10
Doctor	−0.00	0.12	−0.13	0.03	0.14	−0.09
Civil Servant	0.04	0.10	−0.02	−0.03	0.03	−0.10
Local Politician	−0.01	0.23	−0.25	0.10	0.40	−0.21
Business	−0.05	0.21	−0.31	0.00	0.13	−0.13
White Collar	−0.00	0.19	−0.19	−0.00	0.15	−0.16
Union Official	0.00	NA	NA	−0.04	0.12	−0.20
Journalist	−0.08	0.07	−0.22	0.05	0.29	−0.20
Miner	0.00	NA	NA	−0.02	0.02	−0.07
Schooling: Eton	0.12	0.28	−0.04	−0.04	0.02	−0.11
Schooling: Public	−0.22	0.07	−0.52	0.03	0.23	−0.17
Schooling: Regular	−0.15	0.12	−0.42	−0.01	0.32	−0.35
Schooling: Not reported	0.25	0.46	0.03	0.02	0.33	−0.30
University: Oxbridge	0.10	0.36	−0.17	−0.04	0.21	−0.30
University: Degree	−0.02	0.25	−0.30	0.10	0.42	−0.23
University: Not reported	−0.08	0.21	−0.37	−0.06	0.25	−0.37
Aristocrat	0.05	0.19	−0.09	0.06	0.17	−0.06
Vote Margin in Previous Race	−0.00	0.04	−0.05	−0.05	0.01	−0.11
Previous Races	0.22	0.59	−0.16	0.24	0.76	−0.29

Note: Every row shows a placebo treatment effect estimated at the threshold of winning $\tau_{RDD} = E[Y(1) - Y(0)|Z = 0]$ obtained from local linear regression with rectangular kernel (equation 4.2); bandwidth is 15 percentage point of vote share margin. UB and LB refer to the upper and lower bound of the .95 percent confidence interval.

Table 4.6: Characteristics of the Political Careers of Members of Parliament (Estimation Sample)

	Mean	Min	1st. Qu	Median	3rd. Qu	Max
Conservative						
Cabinet	0.13	0	0	0	0	1
Front Bench	0.27	0	0	0	1	1
Year of Birth	1916	1895	1912	1916	1921	1940
Age Entered Office	42	29	37	41	46	59
Year Entered Office	1958	1950	1951	1959	1964	1970
Year Retired from last Office	1977	1955	1966	1974	1987	2001
Years as MP and Former MP	37	14	31	38	45	55
Years served as MP	18	2	9	18	24	51
Years as Former MP	18	0	10	17	28	41
Labour						
Cabinet	0.13	0	0	0	0	1
Front Bench	0.30	0	0	0	1	1
Year of Birth	1920	1901	1915	1920	1926	1935
Age Entered Office	42	31	38	42	46	57
Year Entered Office	1962	1950	1959	1964	1966	1970
Year Retired from last Office	1981	1951	1979	1983	1987	1997
Years as MP and Former MP	34	18	30	34	40	50
Years served as MP	18	1	13	19	24	33
Years as Former MP	15	0	6	14	21	46

Table 4.7: The Correlates of Wealth: Estimates for Conservative MPs (Estimation Sample)

Dependent Variable	Log Wealth	Log Wealth	Log Wealth
Model Number	(1)	(2)	(3)
Years as MP and Former MP	0.029		
	(0.013)		
Years Served as MP		0.033	0.032
		(0.014)	(0.024)
Years as Former MP		0.019	0.017
		(0.014)	(0.023)
Years as MP · Years as Former MP			.0001
			(.0014)
Front Bench or Cabinet	0.36	0.27	0.27
	(0.26)	(0.29)	(0.29)
University: Oxbridge	0.17	0.13	0.14
	(0.28)	(0.27)	(0.26)
University: Degree	0.20	0.16	0.16
	(0.27)	(0.29)	(0.30)
Schooling: Eton	0.91	0.97	0.97
	(0.40)	(0.40)	(0.40)
Schooling: Public School	0.00	0.04	0.04
	(0.28)	(0.27)	(0.28)
Schooling: Regular	0.19	0.14	0.15
	(0.38)	(0.37)	(0.37)
Aristocrat	0.48	0.46	0.44
	(0.33)	(0.34)	(0.36)
Margin of Winning	-1.17	-1.43	-1.46
	(1.52)	(1.56)	(1.55)
Year of Birth	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)
Intercept	-15.94	-3.51	-3.97
	(24.22)	(25.27)	(25.93)

Note: N=104. OLS coefficients with robust standard errors in parentheses.

4.8 Figures for Chapter 4

Figure 4.1: Fraction of Members of Parliament that declared Outside Interests 1975, 1990, and 2007 (fractions by party; dashed (solid) line decodes Labour (Conservatives))

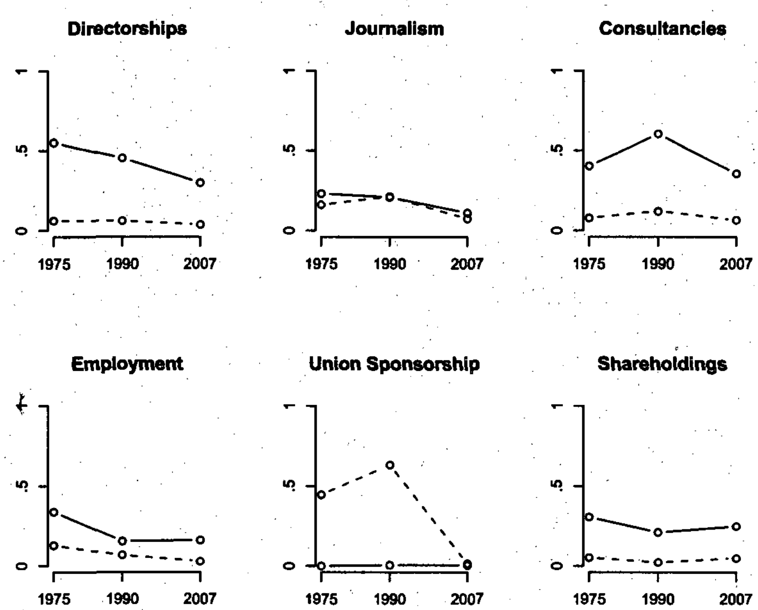
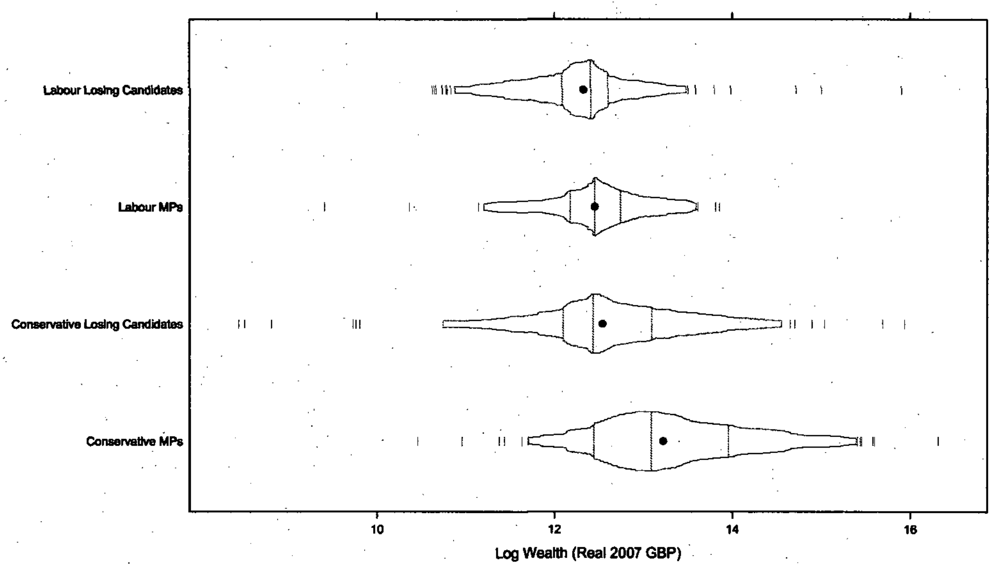
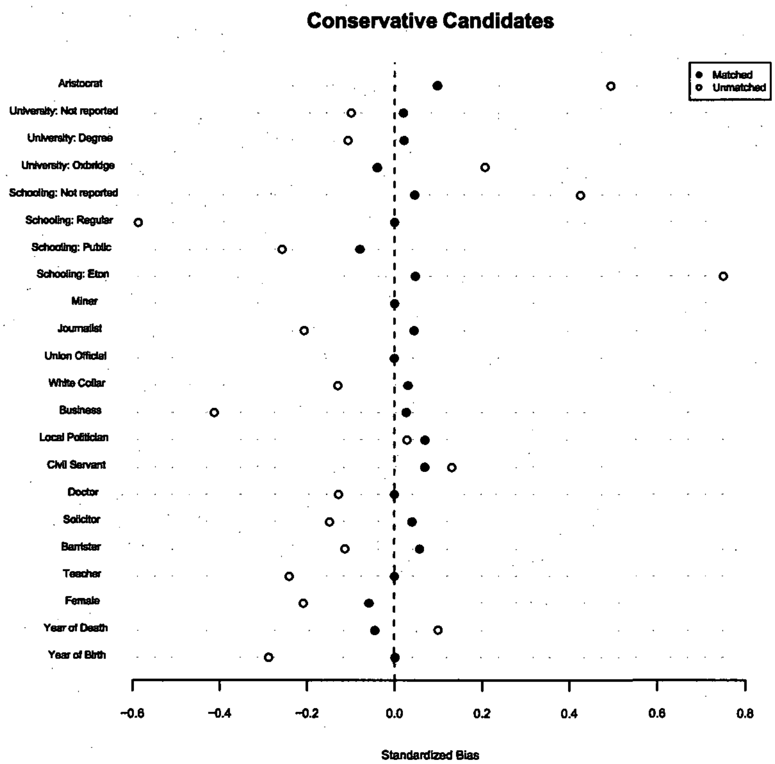


Figure 4.2: Distributions of (Log) Wealth at Death by Party for Winning and Losing Candidates to the House of Commons 1950-1970



Note: Box percentile plots. Box shows empirical distribution function from .05 to .95 quantile; vertical lines indicate the .25, .5, and .75 quantile respectively. Observations outside the .05 – .95 quantile range are marked by vertical whiskers. The dot decodes the mean.

Figure 4.3: Covariate Balance Before and After Matching



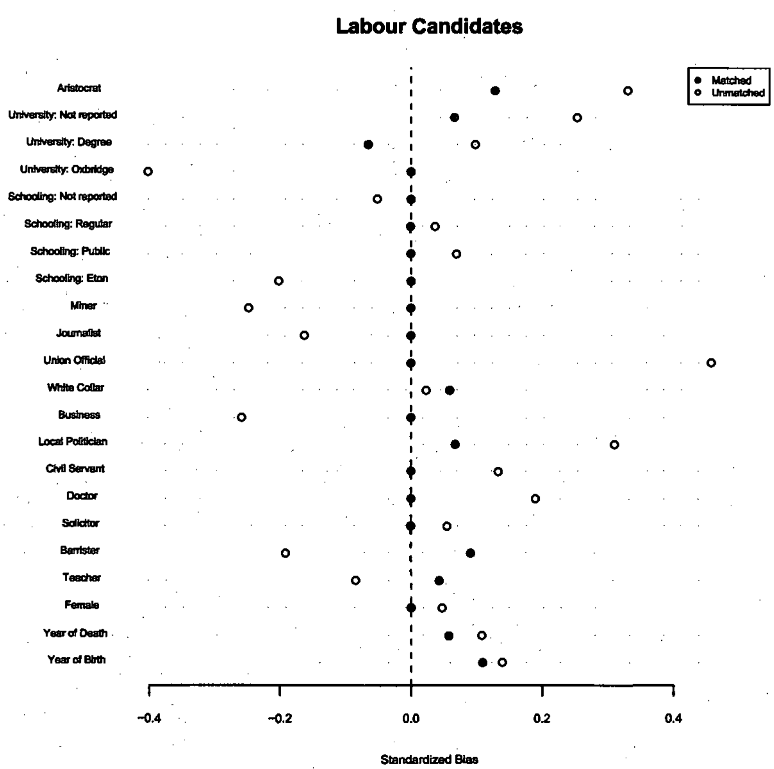


Figure 4.4: Regression Discontinuity Design: The Effect of Serving in the House of Commons on Wealth at Death

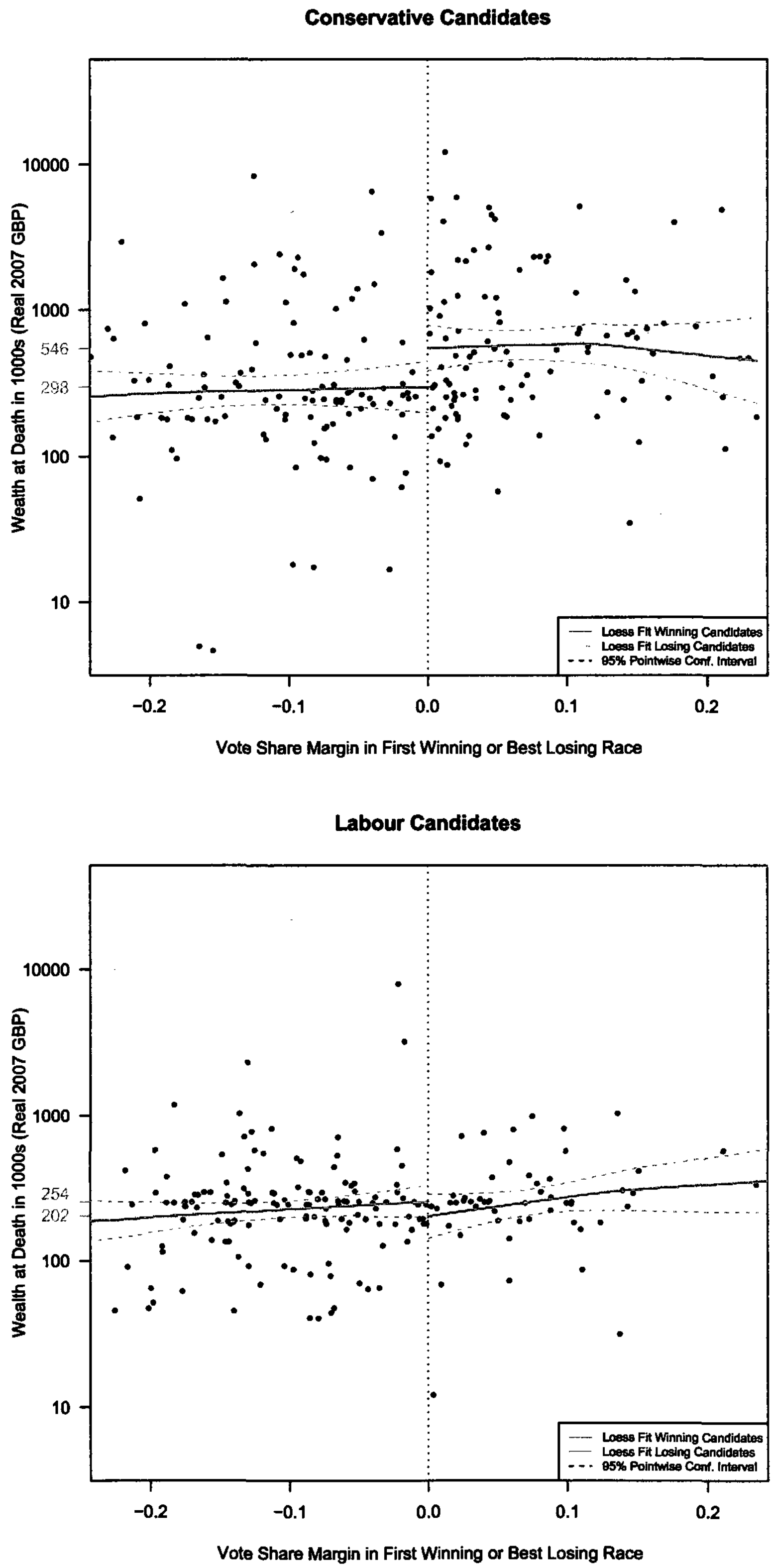
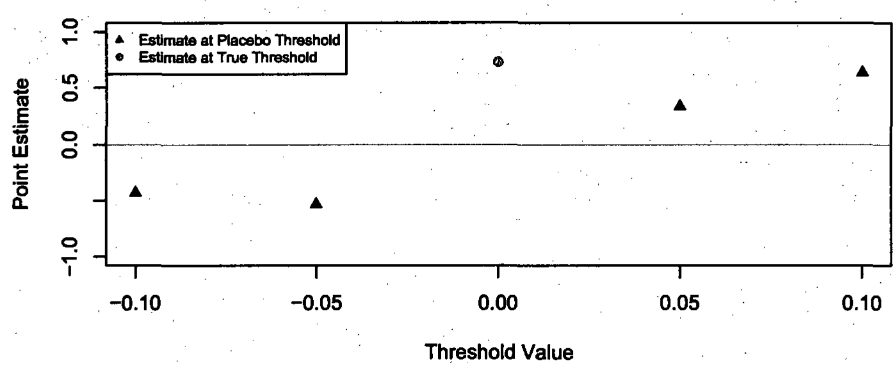


Figure 4.5: Testing for Jumps at Non-discontinuity Points: Estimates for Conservative Candidates



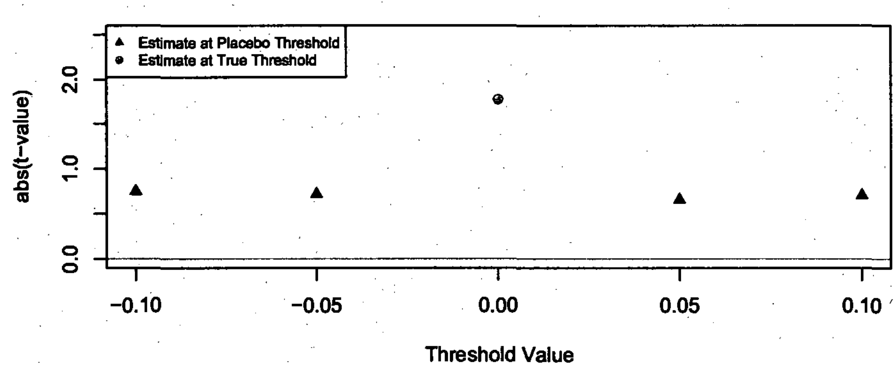
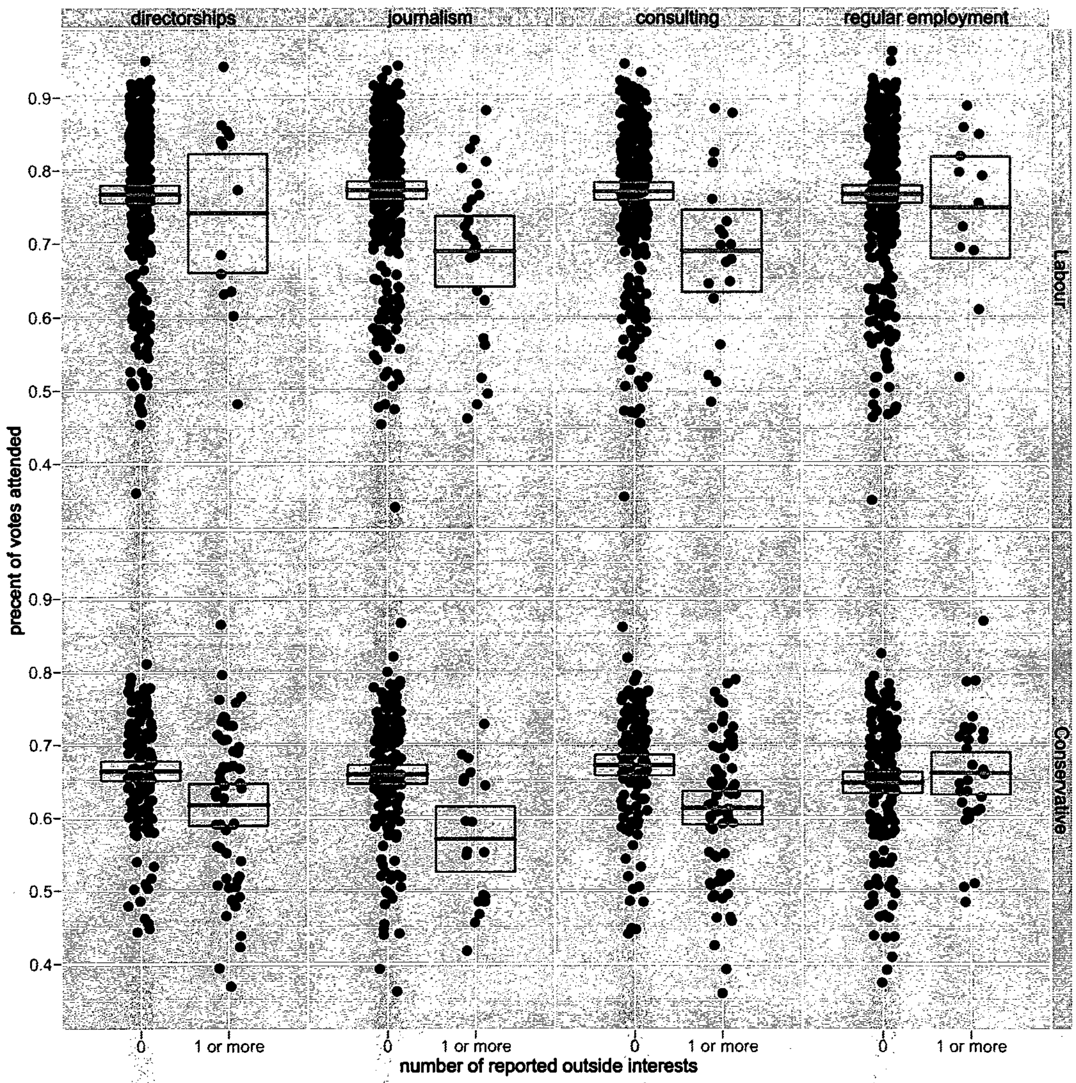


Figure 4.6: Outside Interests and Vote Attendance in the House of Commons 2005-2007



Note: Jittergrams that show the percent of votes attended between 2005-2007 by party and outside interests. Crossbars decode the mean and .95 percent confidence intervals. 326 Labour MPs and 192 Conservative MPs.

Chapter 5

Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes

5.1 Introduction

One of the most prominent subjects in comparative politics during the past few decades has been the explanation of political regime change from authoritarian rule to democracy.[1] Beginning with the pathbreaking work of O'Donnell, Schmitter, and Whitehead (1986), many authors have developed complex theoretical models of the

[1]This paper is co-authored with Holger Kern, Cornell University. We thank Alberto Abadie, Christopher Anderson, Jake Bowers, Daniel Butler, Alexis Diamond, Andy Eggers, Justin Grimmer, Dominik Hangartner, Dan Hopkins, Guido Imbens, Jan Lemnitzer, Walter Mebane, Na'ama Nagar, Beth Simmons, Hans-Jörg Stiehler, Susan Stokes, Christopher Way, Robert Weiner, and seminar participants at Harvard, MIT, Cornell, Dartmouth, Northwestern, UC Davis, and the University of Berne for very helpful comments. Brigitte Freudenberg at the Office of the Federal Commissioner for the Records of the National Security Service of the Former German Democratic Republic was of great help in tracking down archival material. Peter Bischoff, Evelyn Brislinger, Kurt Starke, and especially Hans-Jörg Stiehler patiently answered our many questions about the surveys conducted by the Central Institute for Youth Research. The usual disclaimer applies.

determinants of such transitions. The early literature on democratization concluded that international factors played no more than a secondary role in transitions to democracy, but more recent theoretical and empirical work has put greater emphasis on the international context (Starr 1991; Huntington 1991; Pridham 1991, 1997; Gasiorowski 1995; Linz and Stepan 1996; Whitehead 1996; O'Loughlin et al. 1998; Kopstein and Reilly 2000; Starr and Lindborg 2003; Doorenspleet 2004; Brinks and Coppedge 2006; Gleditsch and Ward 2006). In the wake of this theoretical reorientation, foreign mass media have attracted increased attention as a cause of democratization. The view that Western mass media made a significant contribution to the fall of communism in Eastern Europe is widely shared among scholars and policymakers (Giddens 2000; Nye 2004, 2008; Parta 2007). What makes the wide acceptance of this view puzzling is that empirical research about the impact of foreign mass media on the stability of authoritarian regimes is almost nonexistent. Historical case studies of particular media sources such as *Radio Liberty* and *Radio Free Europe* (Lisann 1975; Nelson 1997; Puddington 2000) are suggestive but do not provide systematic evidence about their impact. A few studies use survey data from authoritarian regimes to document a correlation between exposure to foreign mass media and pro-democratic and pro-Western attitudes but fall short of demonstrating that these correlations are indeed causal (Hesse 1988, 1990; Parta 2007). However, a clear distinction between correlation and causation is particularly crucial in studies of media effects since it is well established that people tend to expose themselves to political messages they agree with (Bartels 1993; Stroud 2008).

In this paper, we address the question of the impact of foreign mass media on public support for authoritarian regimes by exploiting a natural experiment in communist East Germany. Because of East Germany's topography, West German television broadcasts could be received in most but not all parts of the country. We take advantage of this naturally occurring variation to identify the causal effect of West German television exposure on support for the East German regime. We rely on formerly classified survey data, behavioral data on exit visa applications, and archival evidence on the reaction of the East German leadership to the availability of West

German television. Our main result is that exposure to West German television increased support for the East German regime. We argue that the best explanation for this counterintuitive finding is that East Germans used West German television primarily as a source of entertainment. In a society with a very limited number of entertainment options, the ability to watch West German television made a real difference in people's daily lives. It offered them a vicarious escape from the scarcities, the queues, and the ideological indoctrination, and so made life under communism more bearable and the East German regime more tolerable. We do not necessarily argue that West German television's political content did not undermine public support for the East German regime at all. However, the evidence shows that the net effect of West German television exposure was an increase in regime support. Behavioral data on regional patterns in exit visa applications and archival evidence on the reaction of the East German leadership to the availability of West German television corroborate these results.

5.2 The effects of foreign media exposure

During the Cold War, the United States attempted to reach audiences in communist countries through radio stations such as *Voice of America*, *Radio Liberty*, and *Radio Free Europe*. Some of these radio stations were explicitly set up as "surrogate" stations. They specialized in emulating the domestic stations that countries behind the Iron Curtain would have had if they had been free. U.S. international broadcasting policies were based on the expectation that access to uncensored and trustworthy information would nurture a more pro-democratic and pro-Western public opinion, restrain communist militarism and adventurism, and erode public support for communist rule in the long run (Presidential Study Commission on International Radio Broadcasting 1973; Lisann 1975; Quester 1990; Nelson 1997; Puddington 2000).

After the end of the Cold War, both scholars and policymakers agreed that U.S. international broadcasting, together with radio stations sponsored by other Western countries (e.g., *Deutsche Welle* and the *BBC*), had made a significant contribution to the demise of communism in Eastern Europe (Giddens 2000; Nye 2004, 2008; Parta 2007). The democratization literature holds that Western broadcasts sapped the strength of communist regimes by giving people behind the Iron Curtain hope and the assurance that the Free World had not forgotten them. They provided information not available in the state-controlled domestic media, thus allowing Eastern Europeans to compare communist propaganda with credible information from abroad. Western broadcasts addressed issues suppressed in the domestic media such as the existence of dissident movements, human rights violations, and communist countries' involvement in foreign wars. They enabled Eastern Europeans to compare their living standards with the (generally much higher) standards of living in supposedly declining capitalist countries. Through their coverage of domestic politics in democratic nations, they familiarized listeners with the functioning of democracy and introduced them to freedoms and liberties unknown under communism. In the long run, Western broadcasting nurtured pro-democratic attitudes and undermined public support for communism (Rustow 1990; Diamond 1993a, 1993b; Dalton 1994; Rohrschneider 1994, 1996; Whitehead 1996; Bennett 1998; Roberts 1999; Sükösd 2000).

The literature often cites the availability of West German television in East Germany as a particularly vivid example of the democratizing force of foreign mass media. Rustow (1990) for instance notes that West German television invited a constant comparison of living standards between affluent West Germany and its much poorer East German neighbor, thereby undermining the political legitimacy of the East German regime. According to Quester (1990: 128), the availability of West German television posed a "constant threat of destabilization and ferment." Roberts (1999: 33) also draws a direct connection between West German television and the fall of the East German regime, noting that "the availability of West German television in East Germany influenced attitudes over a long period [...], compelling a reluctant regime to initiate change." For Whitehead (1996: 6), "popular attitudes in East Germany were so powerfully influenced by messages transmitted neutrally from the West that democratization became unavoidable, whatever governments or political leaders within or without might have wished or attempted." Rohrschneider (1994: 935) asserts that access to West German television "undoubtedly helped to erode the legitimacy of the East German economic and political system."

West German television was indeed quite popular in East Germany; it was in fact much more popular than East German television (Stiehler 2001).[2] East and West Germans shared the same language and very similar cultures, which should have maximized the potential impact of West German broadcasts.[3] Moreover, West German television devoted a great deal of attention to politics in East Germany (e.g., political magazines such as *Kontraste* and *Kennzeichen D*). There can be no doubt that because of the availability of West German television, most East Germans had access to political information that the East German regime would have preferred them not to have. All in all, the claim that West German television undermined public support for the East German regime is very plausible, but so far it has not been tested empirically.

In its confidence in the power of foreign mass media to undermine authoritarian stability, the democratization literature also fails to consider the possibility that foreign media might inadvertently stabilize authoritarian regimes. West German television not only informed its viewers about the grim reality of communist rule in Eastern Europe, it also threw light on controversial aspects of capitalist societies such as crime, drug abuse, mass unemployment, and the unequal distribution of wealth. German communications scholars have therefore conjectured that West German television had a deterring effect on East Germans, raising their awareness of the "dark sides" of capitalism and increasing their appreciation for the advantages of "real existing socialism" (Hesse 1988: 118–122; Meyen 2003b: 67–68). East Germans could not acquire the consumer goods or travel to the exotic places they encountered on

[2]West German television refers to *ARD* and *ZDF*, the two primary West German public broadcasting stations. Commercial television was introduced in West Germany in the mid-1980s but it did not broadcast to East Germany and could only be received in some areas near West Berlin. Since East Germany bordered Poland and Czechoslovakia, some East Germans could have watched television broadcasts originating in these countries. We feel justified in ignoring this complication here since the Marxist orientation of these stations together with language barriers limited their appeal.

[3]Almost all East German households owned at least one television set in the 1980s (Staatliche Zentralverwaltung für Statistik 1989).

West German television, but they also did not have to worry about unemployment or homelessness. Confronted with social ills that did not exist in East Germany, they might have come to the conclusion that the grass was not much greener on the other side of the Berlin Wall.

Alternatively, people living under authoritarian regimes might simply see the mass media, both domestic and foreign, as a source of entertainment that offers a temporary escape from the humdrum of daily life, the scarcities, the queues, and the ideological indoctrination (Lazarsfeld and Merton 1948; Katz and Foulkes 1962). Especially Meyen (2001, 2003b: 37–42) has stressed the escapism aspect of West German television. As in other communist countries, life in East Germany was rather dull and uneventful. A scarcity of restaurants, cinemas, theaters, and night and sports clubs contributed to domesticity and boosted West German television's importance as the primary source of entertainment. As one German historian has put it, each night East Germans "collectively emigrated" to West Germany in front of their television screens (Wolle 1998: 71). West German television, and in particular its entertainment programming, may have offered East Germans a vicarious escape that made life under communism more bearable and the East German regime more tolerable.

Below, we test the conventional view that West German television exposure undermined support for the East German regime together with the deterrence and escapism hypotheses. The next section of the paper describes our research design and statistical methodology. Section III presents the main results. Section IV shows that access to West German television was also related to the number of exit visa applications filed by East Germans, which serve as a behavioral indicator of political dissatisfaction. Based on research in the archives of the East German secret police and the ruling Socialist Unity Party (SED), section V shows that the East German leadership itself was aware of the stabilizing effects of West German television. The last section discusses our results.

5.3 Research design

One challenge for research on media effects in authoritarian regimes is data availability. Authoritarian regimes rarely permit independent survey research and when they themselves conduct public opinion surveys the results are often not made public. Thus, with some exceptions (e.g., Geddes and Zaller 1989), reliable micro-level data on public opinion in authoritarian regimes are not readily available. Past research on media effects in Eastern Europe has attempted to address this problem by interviewing emigrees or visitors to Western Europe. Hesse (1988, 1990) for example interviewed refugees coming to West Germany. *Radio Liberty* and *Radio Free Europe* interviewed Eastern Europeans visiting the West to estimate the effects of exposure to Western radio programs on political attitudes in Eastern Europe and the Soviet Union (Parta 2007). However, such surveys do not address the problem of self-selection into media exposure. Soviet citizens who listened to *Radio Liberty* almost certainly had different political attitudes than non-listeners to begin with. To attribute differences in political attitudes between these two groups to *Radio Liberty* exposure is therefore problematic.

We rely on formerly classified survey data collected by the Zentralinstitut für Jugendforschung (Central Institute for Youth Research) that have become available to researchers after German re-unification. The Central Institute was founded in 1966 to offer scientific guidance to East German authorities on questions of youth policy. Its critical reports aroused the hostility of parts of the East German bureaucracy and during most of its existence, researchers affiliated with the Central Institute were not allowed to publish any of their research. Between 1966 and 1990, the Central Institute conducted several hundred surveys of East German high school and college students, apprentices, and young workers. For each survey, questionnaires had to be approved by SED party officials, several government departments, and the Central Office for Statistics. Surveys representative of East Germany as a whole could not be conducted for political reasons; the only stratum of the East German population the Central Institute was allowed to survey were teenagers and young adults. The range of permissible survey questions was likewise restricted. Given these constraints, the

Central Institute relied on cluster sampling of teenagers and young adults in schools, universities, and firms (Friedrich, Förster, and Starke 1999).

Researchers at the Central Institute were aware that people living under an authoritarian regime might be reluctant to participate in public opinion surveys and answer politically sensitive questions. In order to convince participants that survey responses would be completely anonymous the researchers eschewed face-to-face interviews. They always distributed questionnaires in group settings (e.g., high school or college lecture classes), thereby giving respondents the opportunity to see for themselves that the questionnaires were unmarked and could not be traced back to them. Completed questionnaires were collected in a sealed urn (Friedrich 1990). Ex post, it is impossible to determine with certainty to what extent these procedures eliminated preference falsification (Kuran 1991). Nonetheless, as we will show below, survey responses are fairly critical of the East German regime. Moreover, even if the existence of preference falsification cannot be completely ruled out it almost certainly does not create a problem for our research design. We are not primarily concerned with identifying the true level of East Germans' regime support. Instead, we estimate the difference in regime support between East Germans who were exposed to West German television and East Germans who were not exposed to West German television. Unless preference falsification varied systematically between these two groups our media effect estimates will be unbiased.

People tend to expose themselves to political messages they agree with, making it difficult to distinguish between association and causation in studies of media effects (Bartels 1993; Stroud 2008). If exposure to West German television is associated with critical attitudes towards the East German regime, it could be that West German television had an effect on viewers' attitudes but it is equally plausible that East Germans dissatisfied with the regime were more likely to watch West German television in the first place. In order to address this self-selection problem we take advantage of a natural experiment. Natural experiments are observational studies in which some exogenous process ("nature") assigns units to different types of treatments in a haphazard fashion that is "as good as random" (Freedman 2005;

Morgan and Winship 2007; Dunning 2008). We identify the effects of exposure to West German television by exploiting a natural experiment generated by East Germany's topography. As it turns out, West German over-the-air television broadcasts could not be received everywhere in East Germany. Especially the Dresden district in the Southeast was largely cut off from West German television due to topographical features and its distance from West German broadcasting towers (Figure 5.1).

5.3.1 Survey data

The survey data we use in the main part of the paper were collected between November 1988 and February 1989, less than one year before the fall of the Berlin Wall in November 1989.[4] It is important to note that at that time, no one expected the demise of the East German regime, which appeared to be one of the most stable communist regimes in Eastern Europe (Zelikow and Rice 1995). This survey is one of the very few surveys conducted by the Central Institute that contain the necessary information for identifying the effect of West German television on regime support (respondents' place of residence, media exposure, baseline characteristics, and attitudinal outcome variables). The sample contains young adults (the average age is 23 years with a standard deviation of 6 years) from 8 East German districts (Dresden, East Berlin, Magdeburg, Cottbus, Leipzig, Erfurt, Karl-Marx-Stadt, and Schwerin). The total number of respondents is $N = 3,564$.

5.3.2 Causal Inference using instrumental variables

Our main quantity of interest is the effect of West German television exposure on expressed support for the East German regime. We code a binary variable D that takes the value 1 for respondents who watch West German television daily, multiple times each week, once per week, or less than once per week, and 0 for respondents

[4]Zentralarchiv für Empirische Sozialforschung ZA 6008: Politisches Klima und gesellschaftliche Bedingungen in der DDR 1988/89.

who never watch West German television.[5] Borrowing from the literature on causal inference in statistics (Rubin 1974, 1978; Holland 1986; Rosenbaum 2002), we refer to D as "treatment indicator," the set of respondents exposed to West German television as "treatment group," and the set of respondents not exposed to West German television as "control group." We define Y_1 and Y_0 as the potential outcomes under treatment and control, i.e., the levels of expressed regime support that a respondent would have had with and without exposure to West German television. For each respondent, the effect of West German television exposure is defined as the difference between these two potential outcomes $(Y_1 - Y_0)$. The "fundamental problem of causal inference" (Holland 1986) is that for each respondent we never get to observe both potential outcomes but only the realized outcome $Y = D \cdot Y_1 + (1 - D) \cdot Y_0$. In other words, for a respondent exposed to West German television we never get to observe the counterfactual level of regime support that she would have had in the absence of exposure to West German television (and vice versa).

In order to obtain unbiased treatment effect estimates we need to find a suitable control group of respondents that is sufficiently similar to the treatment group in all relevant characteristics except that it was not exposed to West German television. Finding a good control group is difficult in observational studies because selection into treatment is usually associated with the potential outcomes. It is for this reason that the naive comparison of the regime support of East Germans exposed to West German television with the regime support of East Germans not exposed to West German television would not be very informative about the causal effect of West German television exposure. We could attempt to make this comparison more plausible by

[5]Respondents were asked "How often do you watch West German television?" Responses are coded on a five-point scale ranging from daily to never. We dichotomize this variable to facilitate the interpretation of our results. One could also think of West German television exposure as a multi-valued treatment but that would make identification much more complicated (see Wooldridge and Imbens 2008 for details). Moreover, note that East Germany's topography induced variation in whether East Germans watched West German television and not the specific amount they watched. In other words, West German television exposure might only be "as good as random" when dichotomized. In any case, results are substantively identical when respondents watching West German television less than once per week are also coded as 0 or when we keep all 5 categories and add a linearity assumption.

controlling for observable characteristics known to affect both selection into West German television exposure and political attitudes using some method of covariate adjustment (regression or matching, for example). Such an approach would still be problematic, however, since we cannot rule out the possibility that East Germans selected into West German television exposure based on unobservable characteristics.

Instrumental variables (IV) (Imbens and Angrist 1994; Angrist, Imbens, and Rubin 1996; Abadie 2003; Wooldridge and Imbens 2008) offer a more credible identification strategy which exploits the fact that not all parts of East Germany had equal access to West German television. We code a binary instrument Z based on information about respondents' place of work. It takes the value 0 for respondents living in the Dresden district and the value 1 for respondents living in districts with access to West German television (East Berlin, Magdeburg, Cottbus, Leipzig, Erfurt, Karl-Marx-Stadt, and Schwerin). One can think of this instrument as inducing exogenous variation in the treatment.[6]

We follow Imbens and Angrist (1994) in conceptualizing IV identification in terms of potential treatment indicators. Let D_z represent potential treatment status given $Z = z$. For example, $D_0 = 0$ and $D_1 = 1$ means that a respondent would not watch West German television if she lived in the Dresden district but that she would watch West German television if she did not live in the Dresden district. The treatment status indicator can then be expressed as $D = Z \cdot D_1 + (1 - Z) \cdot D_0$. Similar to the missing data problem for potential outcomes, we only get to observe Z and D (and therefore D_z for individuals with $Z = z$) but never both potential treatment indicators for the same individual. Following the terminology in Angrist, Imbens, and Rubin (1996) we can distinguish 4 groups of respondents:

Compliers: $D_0 = 0$ and $D_1 = 1$. Compliers are respondents who watch West German television if they do not live in Dresden but who do not watch West German

[6]Could we use a regression-discontinuity design (RDD) instead? RDDs are sometimes used with spatial treatment discontinuities (e.g., Black 1999). However, our survey data do not allow us to implement a spatial RDD because doing so would require information about respondents' place of residence at a much more disaggregate level (so that we could code a running variable that measures their precise distance from the boundary that divides the parts of East Germany with access to West German television from the parts of East Germany without access to West German television).

television if they live in Dresden.

Always-takers: $D_1 = D_0 = 1$. These are respondents who always watch West German television no matter where they live.

Never-takers: $D_1 = D_0 = 0$. Similarly, these are respondents who never watch West German television no matter where they live.

Defiers: $D_0 = 1$ and $D_1 = 0$. These are respondents who watch West German television if they live in Dresden but who do not watch West German television if they do not live in Dresden.

Since we only observe one of the potential treatment indicators (D_0, D_1) we cannot directly identify the group to which any particular respondent belongs. However, it turns out that under certain assumptions that we will discuss next, instrumental variables allow us to estimate average treatment effects for the subgroup of compliers (62 % of our respondents) even though we cannot individually identify compliers in our sample.

Identification assumptions

Let Y_{zd} represent the potential outcome if $Z = z$ and $D = d$. For example, Y_{00} denotes the regime support of a respondent who lives in the Dresden district and does not watch West German television. Following Abadie (2003), this leads to the following nonparametric assumptions under which instrumental variables can be used to identify causal effects. X represents a vector of predetermined covariates.

(i) Independence of the instrument: Conditional on X, the random vector $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, D_0, D_1)$ is independent of Z.
(ii) Exclusion of the instrument: $P(Y_{1d} = Y_{0d} \mid X) = 1$ for $D \in \{0, 1\}$.
(iii) First stage: $0 < P(Z = 1 \mid X) < 1$ and $P(D_1 = 1 \mid X) > P(D_0 = 1 \mid X)$.
(iv) Monotonicity: $P(D_1 \geq D_0 \mid X) = 1$.

Are these identification assumptions plausible? Assumption (iv) rules out the existence of defiers. It is highly unlikely that there were East Germans who would have watched West German television if they had lived in Dresden but who would not have watched West German television if they had not lived in Dresden. Assumption (iii) is also innocuous. It guarantees that Z and D are correlated conditional on X. Given that most people in the Dresden district were cut off from West German television broadcasts, not living in the Dresden district is highly correlated with exposure to West German television. Table 5.1 displays the frequency of West German television consumption in our sample. As one would expect, respondents living in the Dresden district were much less likely to watch West German television than respondents living in other districts. The sample correlation between living in Dresden (Z) and consumption of West German television (D) is 0.74. When we regress Z on D while controlling for our extensive set of covariates (see below) the t-statistic on the treatment indicator is about 60.

Assumptions (i) and (ii) need more justification. Assumption (i) states that place of residence (Z) is "as good as randomly assigned" once we condition on X. Assumption (ii) implies that variation in the instrument does not change the potential outcomes other than through its effect on D. It allows us to define potential outcomes in terms of D alone, so that we have $Y_0 = Y_{00} = Y_{10}$ and $Y_1 = Y_{01} = Y_{11}$. Taken together, these two assumptions guarantee that conditional on X the instrument only has an effect on the outcome through the variation it induces in the treatment. Conditional on covariates, living in Dresden as such does not directly affect respondents' political attitudes.

Assumptions (i) and (ii) might only hold when we condition on a set of covariates X. Our research design enables us to control for confounding factors at two levels. Below, we show that the Dresden district was very similar to the other East German districts in our sample. In this sense, aggregate-level differences are controlled for by design. In addition, the survey contains a number of individual-level characteristics that we can directly control for by including them in the estimations.

Threats to the exclusion restriction

The exclusion restriction would be called into question if the Dresden district was fundamentally different from other districts in our sample. Living conditions in Dresden for example might have been worse than in other parts of East Germany, making Dresden respondents more likely to express dissatisfaction with the East German regime.

Figure 5.2 shows that this possibility is rejected by the data. It displays characteristics for the 8 districts in our sample for the year 1988 (Staatliche Zentralverwaltung für Statistik 1989; Grundmann 1997). The x-axis shows deviations from the medians; the y-axis lists the district characteristics. Red solid circles display values for the Dresden district; empty circles display values for the other districts. The first 21 rows in Figure 5.2 list socio-economic characteristics; they do not provide any indication that the Dresden district is fundamentally different from the other districts.

What about differences in political attitudes before West German television ever became available in East Germany? The last 4 rows of Figure 5.2 compare turnout and vote shares for the Christian Conservatives (CDU), Liberals (LDP), and the SED in the 1946 state elections held in the German territories then occupied by the Soviet Union.[7] These elections were the first state elections after the collapse of the Third Reich. Although the Soviet occupation forces supported the SED and discriminated against the CDU and LDP the elections were still reasonably free (Hajna 2000; Schmitt 1993). As we can see in the second to last row in Figure 5.2, the SED's vote share in the Dresden district was close to average. The same is true for turnout and support for the CDU and LDP. There is no evidence that long-standing regional differences in political culture predating the introduction of West German television invalidate the exclusion restriction.

[7]The data are taken from Broszat and Weber (1993), adjusted for redistricting, and weighted by population size. East Berlin has been excluded since West German parties could run their own candidates in East Berlin at that time.

Threats to the ignorability assumption

East Germans who desired to watch West German television might have moved away from the Dresden district. If interest in West German television was correlated with regime support, which seems likely, this kind of sorting behavior would invalidate our instrument. Residential mobility however was exceptionally low in East Germany. On average, East Germans moved across county borders once every 64 years, i.e., approximately once per lifetime (Grundmann 1998).[8] The lack of functioning labor and housing markets accounts for this extremely low level of residential mobility. In East Germany's centrally planned economy, factor allocation was controlled by the state. Mobility of labor between occupations and firms was not desired since it interfered with central planning and increased costs, and low labor mobility lead to low residential mobility. Second, a dramatic shortage of housing depressed residential mobility further (Wolle 1998: 182–88; Grundmann 1998; Grünert, Bernien, and Lutz 1997; Grünert 1996; Uunk, Mach, and Mayer 2005). We therefore think it highly unlikely that the ignorability assumption is violated by spatial sorting.

5.3.3 Covariates

Even though the use of a natural experiment reduces the danger of confounding, some imbalances between the treatment and control groups might exist. Fortunately, the survey contains a relatively rich set of control variables that allows us to relax the exclusion restriction for the instrument. Here we face the usual tradeoff involved in the choice of covariates (Rosenbaum 2002: 76). We want to avoid post-treatment bias, which is caused by adjusting for variables that are themselves affected by the instrument or the treatment (Rosenbaum 1984). Since we only have cross-sectional data at our disposal it is not immediately obvious which variables are pre-treatment in this sense. Age, gender, and father's and mother's occupational classification are the only variables in the survey that are clearly causally prior to television exposure.

[8]East Germany was divided into 15 districts (including East Berlin), which in turn were divided into counties. The total number of counties was 217.

We call this set of variables the limited covariate set. But we also want to maximize the credibility of the ignorability assumption. We therefore also show results for an extensive set of covariates. This set adds respondent's marriage status, living situation (whether she is living alone or with a partner), number of children, highest educational attainment, occupational classification, net monthly income, and employment status to the limited covariate set. For some of these variables it is not entirely inconceivable that they might be affected by the instrument or the treatment. As we will see, however, our results are virtually identical no matter which covariate set we use. Except for age (which is discretized into 8 categories), all covariates are fully factorized to avoid functional form assumptions. Missing data, which affect less than 1% of our sample, are treated as additional categories.

5.3.4 Estimators

In the absence of covariates the so-called Wald estimator identifies average treatment effects for compliers, also called local average treatment effects (LATE) (Imbens and Angrist 1994):

$$\alpha_{LATE} = \frac{\text{cov}(Y,Z)}{\text{cov}(D,Z)} = \frac{\mathbb{E}[Y \mid Z=1] - \mathbb{E}[Y \mid Z=0]}{\mathbb{E}[D \mid Z=1] - \mathbb{E}[D \mid Z=0]} = \mathbb{E}[Y_1 - Y_0 \mid D_1 > D_0] \quad (5.1)$$

Recall that compliers are individuals for which treatment status is exogenously manipulated by the instrument. In the absence of covariates, the popular two-stage least squares estimator (2SLS) reduces to the Wald estimator and thus also identifies LATE. Once we condition on covariates, however, 2SLS no longer identifies LATE unless treatment effects are constant within strata of X (Abadie 2003; Morgan and Winship 2007). This constant treatment effects assumption is often implausible since it requires that the treatment has the same effect for individuals with the same covariate values. In particular, the treatment effect for compliers is not allowed to differ from the treatment effect for non-compliers. In order to allow for heterogeneous treatment effects we also implement a new class of instrumental variable estimators called local average response functions (LARF) that has recently been proposed by Abadie

(2003). The semi-parametric LARF estimator allows for the identification of LATE conditional on covariates even without the constant treatment effects assumption.[9]

5.3.5 Outcome variables

The survey includes several questions that measure support for the East German regime. Here we focus on the three questions that appear to be most relevant:[10]

To what extend do you agree with the following statements:

- *"I am convinced of the Leninist/Marxist worldview."*
- *"I feel closely attached to East Germany."*
- *"In East Germany, political power is exercised in ways consistent with my views."*

Response categories are fully disagree, largely disagree, largely agree, and fully agree. Descriptive statistics for all three outcome variables are shown in Table 5.2.

5.4 Results

Table 5.3 presents the first set of results. It displays treatment effect estimates for exposure to West German television for four estimators.

The first column of Table 5.3 shows estimates for the naive difference in means between exposed and unexposed respondents. Results for all three outcome variables

[9]Standard errors are adjusted for the potential non-independence of observations within clusters (i.e., the classrooms from which respondents were randomly sampled). There are 169 clusters in our sample. For the 2SLS models, potential within-cluster correlation is accounted for by the Eicker-Huber-White sandwich estimator. For the LARF estimator no closed-form solution for a robust variance estimator exists. Instead, we rely on a cluster-adjusted pairs bootstrap with clusters randomly drawn with replacement (Davison and Hinkley 1997: 101–103). 2,000 bootstrap samples are used for standard errors and 15,000 for confidence intervals.

[10]Results are substantively identical when other questions are used instead.

suggest that watching West German television had only a small and statistically insignificant effect on regime support. However, there is every reason to believe that these estimates are confounded by self-selection. They are not informative about the causal effect of West German television exposure on regime support.

Column 2 displays unconditional LATE estimates. According to these specifications, exposure to West German television increased regime support by $0.15 - 0.22$ on the 4-point response scale. These treatment effect estimates are statistically significant at the .10 level or better.

The unconditional LATE estimates are confirmed when we introduce respondent-level covariates in columns 3 to 6. Exposure to West German television continues to have a positive effect on regime support; effect sizes vary between 0.19 and 0.26 depending on the exact specification. All of these estimates except one are statistically significant at the .05 level. Estimates for the sub-group of compliers (LARF, columns 4 and 6) and the whole sample (2SLS, columns 3 and 5) are virtually identical, which suggests that the constant treatment effects assumption is not critical in our application. The limited and extensive covariate sets also lead to very similar results. Overall, we find that exposure to West German television had a substantial positive effect on regime support. Evaluated at the means of the three outcome variables, the average increase in regime support caused by West German television exposure is between 11 and 15 percent.[11]

How can we explain this surprising positive effect of West German television exposure? The survey also contains several questions that allow us to test the deterrence and escapism hypotheses. Recall that both hypotheses predict that exposure to West German television increased support for the East German regime. The deterrence hypothesis posits that West German television exposed East Germans to negative aspects of West German society and thereby convinced them of the dangers of capitalism and the merits of communist rule. The escapism hypothesis in contrast predicts that West German television, and in particular its entertainment programming, al-

[11]LARF ordered probit results, shown in the online supplement, are substantively identical to the results from these linear models.

lowed East Germans to vicariously escape the dull life under communism at least for a couple of hours each night, which made their lives more bearable and the East German regime more tolerable.

One of the survey questions asked respondents to what extent they felt attached to West Germany (upper half of Table 5.4).[12] If the deterrence hypothesis is correct and West German television exposed East Germans to the "dark side" of West German society, we would expect to see a negative effect of West German television exposure on attachment to West Germany. And this is indeed what we find, although the effect estimate is not all that large and does not quite reach conventional levels of statistical significance. Exposure to West German television on average decreased attachment to West Germany by about .17 on the 5-point response scale. This result is consistent with the deterrence hypothesis, although we are the first to admit that it would be desirable to have survey items on specific aspects of West German society instead of this more general question. Political constraints prevented the Central Institute from adding detailed questions about West Germany to their questionnaires.

The escapism hypothesis receives much stronger support from the data. The survey also asked respondents about their satisfaction with options for recreational activities. If West German television was mainly a means of vicarious escape from "real existing socialism," exposure to West German television should have increased respondents' satisfaction with how they could spend their free time. As one can see from the lower half of Table 5.4, exposure to West German television indeed had such an effect. On average, it increased satisfaction by about 0.40 on the 5-point response scale. This effect is more than twice as large as the deterrence effect; it is also highly statistically significant. Even if deterrence contributed to the positive effect of West German television exposure on regime support, the primary causal channel seems to be escapism.

These results are consistent with previous findings on the use of television in East Germany. We know from detailed survey data on East Germans' television viewing

[12]Note that only a random subset of survey participants was asked these questions. The sample sizes in Table 5.4 are therefore smaller than in Table 5.3.

habits that they watched television primarily for its entertainment value.[13] When asked what changes they would like to see in East German television programming they voted for more entertainment and less politics (Meyen 2001, 2003a). Former East Germans' television viewing habits still differ from those of West Germans although all Germans now have access to the same channels. Former East Germans have a stronger preference for entertainment programming than West Germans. They also favor commercial stations, which put greater emphasis on infotainment and soap operas, over public broadcasting stations (Früh and Stiehler 2002; Darschin and Zubayr 2000).

We have conducted extensive robustness checks, all of which are fully documented in the online supplement. We have replicated our analysis for regional sub-samples, each composed of respondents living in Dresden and in one of the 7 districts with access to West German television. If there was regional variation in unobserved confounders such as respondents' levels of preference falsification, there should be heterogeneity in treatment effect estimates across these sub-samples. There is none. We have also replicated our analysis using data from an earlier survey conducted in 1984. In addition to respondents from Dresden, it contains a sample of young adults from several smaller East German cities (Halle, Schwerin, Dessau, Magdeburg, Bitterfeld). Across all specifications, West German television exposure again has a positive effect on East Germans' regime support, with effect sizes similar to the ones reported here. The survey also contains 167 respondents from Greifswald, a town in the Northeast of East Germany that for the most part also had no access to West German television (see Figure 5.1). We re-estimated our models using these Greifswald respondents as control group (instead of Dresden respondents) and again found a positive effect of West German television exposure. Taken together, these additional survey data results demonstrate that the findings reported here are not driven by biases particular to the 1988/89 survey or the use of Dresden respondents as a control group.

[13]Ergebnisse der Programmarbeit im Zeitraum 1. Juli 1988 bis 8. Januar 1989, Deutsches Rundfunkarchiv, Programmredaktion Analyse II/1988.

5.5 Exit visa applications

Is it possible to corroborate our survey data results with other indicators of regime support? A very direct measure of political (dis-)satisfaction is the decision to emigrate. Even though East Germans had no legal right to emigrate to another country some East Germans insisted on moving to West Germany. Filing an application for such an exit visa had severe consequences. Applicants faced numerous punishments such as getting fired from their jobs, informal discrimination at work, confiscation of their passports, and criminal persecution (Mayer 2002: 177–181; Pfaff 2006).

For the districts in our sample, Table 5.5 shows the number of East Germans per 1,000 residents that filed a new exit visa application in 1988. In the Dresden district, the rate at which such applications were filed was more than twice as high as the average.[14]

While Table 5.5 is suggestive, it is preferable to examine the relationship between exit visa applications and access to West German television at a more disaggregate level. In the mid-1980s, the Department of Postal Services, responsible for radio and television broadcasting in East Germany, conducted an investigation into West German television's over-the-air signal strength in the 17 counties that made up the Dresden district.[15] The resulting map, submitted to the Politburo as part of a larger collection of material on West German television in East Germany, is shown as Figure 5.3. It shows counties without access to West German television (white), counties with partial access to West German television, depending on the weather and exact location (light grey), and counties with access to West German television (dark grey).

We were also able to obtain county-level exit visa application data for 1985–1988.[16] If exposure to West German television had a positive effect on regime support, we would expect the rate at which new exit visa applications were filed to be negatively

[14]The same pattern existed throughout the 1980s. Official statistics can be found in Office of the Federal Commissioner for the Records of the National Security Service of the Former German Democratic Republic (hereafter BStU) MfS-ZKG 10734.

[15]Federal Archive, Foundation for the Archive of the Parties and Mass Organizations of the GDR (hereafter BArch SAPMO) DY 30/J IV2/2/2317.

[16]BStU MfS BV Dresden Stellvert. OP 21; BStU MfS BV Dresden BKG-8028/8029.

associated with the extent to which residents of these counties had access to West German television. For each year, Figure 5.4 plots the distribution of exit visa application rates for counties with access, partial access, and no access to West German television. It displays a clear dose-response relationship: In all four years, the rate at which exit visa applications were filed was highest in the group of counties without access to West German television and lowest in the group of counties with access to West German television. Rates in the group of counties with partial access were always in-between. These differences are large in substantive terms. Averaging over all four cross-sections, the mean rate at which applications were filed in counties without access to West German television was more than twice as high as in counties with access to West German television (6.03 versus 2.82 applications per 1,000 residents; $p = 0.02$).

5.6 Historical evidence

A third type of empirical evidence that we can utilize is archival material on the reaction of the East German regime to the availability of West German television. We expect that if West German television helped to stabilize the East German regime, this surprising fact would not have gone unnoticed by East Germany's massive state security apparatus and the East German leadership.

Already in the 1960s, a large majority of East Germans, including many SED members, watched West German television. Although it was never illegal to watch West German television (except for members of the armed forces and the police), in the 1960s and 1970s East German authorities attempted to suppress its popularity through propaganda and harassment campaigns. The East German regime considered West German television a source of ideological destabilization and feared its effects on public opinion. It therefore attempted to restrict access as far as possible.[17] After the construction of the Berlin Wall in 1961, for example, SED youth squads tore down

[17] BArch SAPMO 30/IVA2/902/68.

hundreds of roof antennas directed towards West Germany. The extreme unpopularity of such campaigns however meant that they were soon abandoned (Holzweissig 2002: 49–65; Stiehler 2001: 13–16; Wolle 1998: 69–71).

The attitude of the East German regime eventually changed. It had no choice but to accept that a large majority (85–90%) of East Germans regularly tuned in to West German television. By 1984, 40% of East German households were connected to community antennas, large antenna systems connected to individual households via cable. Community antennas often carried West German channels.[18] East Germans not connected to community antennas relied on smaller roof antennas. In some areas, however, reception of West German broadcasts was poor or even impossible and residents increasingly complained to state authorities about their inability to watch "international programs," as West German television was euphemistically called.[19] State-owned housing companies sometimes paid for the installation of community antennas, thus providing their tenants with access to West German television.[20] In other cases, East Germans joined forces in grass-roots efforts and paid for the installation of community antennas themselves. While East German authorities were suspicious of any such grass-roots activities and kept a close eye on them, they did nothing to prevent them (Stiehler 2001: 95–96).

Since the mid-1980s, East German authorities were also increasingly confronted with requests from individuals and community initiatives to permit the installation of satellite dishes. Such requests came especially from parts of East Germany where West German over-the-air broadcasts could not even be received with community antennas (such as parts of the Dresden district). The private use of satellite dishes was officially prohibited but some East Germans nonetheless installed satellite dishes they had illegally imported from West Germany or Hungary.[21] When the Department of Postal Services, which was responsible for the enforcement of telecommunications

[18]BArch SAPMO DY 30/IV2/2.039/276: 4–12.

[19]BStU 3189/87 IV: 2–5; BStU 3189/87 III: 58–61, 156–7.

[20]BStU 3189/87 IV: 6, 62.

[21]BArch SAPMO DY 30/IV2/2.039/276: 36–40, 47–49.

regulations, demanded the removal of such satellite dishes and imposed fines their owners would simply remove them only to re-install them later.[22]

In a March 1988 report to the Politburo, the Department conveyed a dramatic picture of the prevailing public mood: "Demands from grass-roots initiatives and individual citizens to allow the reception of West German television [via satellite dishes] are becoming more and more brazen. There are constant disputes. The whole topic is increasingly becoming a political issue, especially since these grass-roots initiatives and individual citizens attempt to gain access to satellite television at all costs and fines imposed for violations of legal regulations do not have the desired effect."[23] Several months earlier, the Department had insisted on the removal of a satellite dish in Marienberg that provided 4,000 households with access to West German television. This ruling led to massive protests and even local SED party officials recommended that the satellite dish should be allowed to stay since its removal would "lead to serious confrontations with thousands of citizens."[24] A similar case happened in the town of Weissenberg. Residents of Weissenberg too had illegally installed a satellite dish. Pleading with state authorities who insisted on its removal, local SED party officials as well as the mayor pointed out that members of their community were "much more content" since the introduction of West German television. Their attitudes towards the East German regime had become "more positive" and all applications for exit visas had been withdrawn. The dismantling of the satellite dish would cause "enormous political problems" since it would contradict the wishes of the entire town.[25]

The Politburo decided in August 1988 on the basis of such reports to almost completely liberalize the private use of satellite dishes (Holzweissig 2002; Stiehler 2001: 95).[26] Six months later, in February 1989, 203 satellite dishes were officially registered

[22]BArch SAPMO DY 30/IV2/2.039/276: 40–43.

[23]BArch SAPMO DY 30/IV2/2.039/276: 37.

[24]BArch SAPMO DY 30/IV2/2.039/276: 41.

[25]BArch SAPMO DY 30/IV2/2.039/276: 45.

[26]BStU MfS BV Dresden 11147.

in the Dresden district, with 250 further applications pending.[27] The archival evidence we have summarized here illustrates a dramatic change in the regime's attitude towards West German television. In the 1960s and 1970s, West German television was seen as undermining public support for communist rule. But during the 1980s, the East German regime began to realize that West German television, not unlike ancient Rome's "bread and circuses," helped to keep its citizens content and docile. In 1988, the East German leadership apparently even contemplated building broadcast relay stations that would have amplified the West German television signal so as to extend its range to the Dresden district (Wolle 1998: 69–71).

5.7 Discussion

In this paper, we have taken advantage of a natural experiment in East Germany to test the commonly held view that foreign mass media undermine the stability of authoritarian regimes. We have found that exposure to West German television increased support for the communist regime among East German young adults. Our analysis suggests that the best explanation for this counterintuitive finding is escapism: West German television, and especially its entertainment programming, allowed East Germans to vicariously escape life under communism at least for a couple of hours each night, which made their lives more bearable and the East German regime more tolerable. West German television also broadcast high-quality hard news programs, some of which offered in-depth coverage of politics in East Germany. There is no doubt that this political content could have undermined public support for the East German regime. What we have found, however, is that West German television exposure resulted in a net increase in regime support.

Behavioral data on exit visa applications and archival material on the reaction of the East German leadership to the availability of West German television corroborate

[27] BArch SAPMO DY 30/J IV2/2/2317: 72–77. Note that satellite dishes were still not available in East Germany and had to be imported from abroad, which prevented most East Germans from obtaining one.

our results. We have found that in the Dresden district, where levels of access to over-the-air West German television broadcasts varied from county to county, exit visa application rates were systematically higher in counties without West German television. The archival evidence clearly shows that by the late 1980s, East German authorities had come to realize that West German television was contributing to the stability of the regime, not undermining it. In an ironic twist for Marxism, capitalist television seems to have performed the same narcotizing function in communist East Germany that Karl Marx had attributed to religious beliefs in capitalist societies when he condemned religion as "opium of the people."

Our survey only includes young adults, which raises questions about the generalizability of our findings to other strata of the East German population. The fact that the behavioral and archival evidence agrees with our survey data results suggests that our inferences are not necessarily limited to young adults. But given the lack of public opinion data representative of East Germany as a whole, we cannot prove that West German television exposure had the same effect on other strata of the East German population.

We are the first to admit that our results do not provide a comprehensive explanation for why some authoritarian regimes are able to stay in power while others are replaced by democratic governments. The study of individuals' political preferences, including how they are shaped by exposure to foreign mass media, is only one step towards a better understanding of authoritarian stability and decline (Welzel 2006; Welzel and Inglehart 2008). Work complementary to ours has analyzed the conditions under which individuals' political preferences lead to collective revolutionary action (Granovetter 1978; Bikhchandani, Hirshleifer, and Welch 1988; Kuran 1991; Lohmann 1994) and how the international opportunity structure affects the emergence and success of popular movements against authoritarian rule (Pridham 1991, 1997; Whitehead 1996; Linz and Stepan 1996). Even though our results are context-specific (East Germans in the middle to late 1980s) and limited to the impact of one specific foreign media source, we believe they usefully complement the existing literatures on transitions to democracy and media effects and political legitimacy in authoritarian regimes.

We believe our results also have relevance for contemporary U.S. international broadcasting efforts. After the attacks of September 11, Congress substantially increased funding for public diplomacy activities, which had been significantly reduced after the end of the Cold War (Nye 2004, 2008). In 2008 alone, the United States has spent almost \$700 million on radio and television broadcasts to the Middle East (*Radio Sawa* and *Alhurra TV*), Eastern Europe (*Radio Free Europe/Radio Liberty*), Asia (*Radio Free Asia*), Cuba (*Radio/TV Martí*), and the *Voice of America.* As during the Cold War, the goal of contemporary U.S. international broadcasting is to win the hearts and minds of foreign audiences, to increase understanding of and support for U.S. foreign policies, and to shift public opinion in non-democratic regimes in a more pro-democratic direction. But if history is any guide, foreign audiences might be more interested in the entertainment value of U.S. broadcasts than their political content.

Finally, our results also offer a caveat to optimistic predictions about the democratizing force of modern information technologies such as the internet. Authoritarian regimes seem to be well aware of the fact that they not only break down national barriers to the transmission of information but also provide entertainment. China for instance successfully pursues a strategy of access *and* control, embracing the economic and entertainment uses of the internet to promote development and sustain public support while strictly controlling political content (Kalanthil and Boas 2003; Shirk 2007).

5.8 Robustness Supplement

This supplement contains the results of all additional analyses we have conducted. For space reasons, we have only summarized these results in the main text of the chapter.

5.8.1 LARF ordered probit

Table 5.6 presents LARF estimates based on an ordered probit link function that takes the ordered categorical nature of the outcome variables into account. We display simulated first differences instead of hard-to-interpret ordered probit coefficient estimates. The local average treatment effect estimate is the average change for compliers in the predicted probability of choosing each of the four response categories. Model-based imputation is used to impute missing potential outcomes under treatment and control, with covariates held at their observed values.

The ordered probit results are qualitatively similar to the results from the linear specifications presented in Table 3 in the paper. For all three outcome variables, West German television exposure on average increases (decreases) the probability of fully agreeing (fully disagreeing) with the survey questions by about 0.06 to 0.13 (0.03 to 0.04).

5.8.2 Subsample comparison

We have checked the robustness of the results presented in Table 3 in the paper by re-estimating our specifications for six sub-samples, each composed of respondents living in Dresden and in one of the districts with access to West German television (East Berlin, Erfurt, Karl-Marx-Stadt, Magdeburg, Leipzig, and Schwerin).[28] We use this sub-group analysis to capitalize on potential variation in unobserved confounders associated with these districts. If there was regional variation in unobserved

[28]We had to exclude the Cottbus district from this analysis because only 66 respondents in our sample reside there.

confounders such as respondents' levels of preference falsification, there should be heterogeneity in treatment effect estimates across these sub-samples.

Table 5.7 demonstrates that this is not the case for any of the districts except East Berlin.[29] It shows linear LARF treatment effect estimates (using the limited covariate set) for each sub-sample. Across sub-samples, treatment effect estimates are similar in effect size and statistical significance. They are also close to the estimates obtained for the whole sample (Table 3 in the paper). This stability in estimates confirms that our finding of a positive effect of West German television exposure is not driven by regionally varying confounders.

5.8.3 Replication I: Alternative survey

Replication is as important in observational studies as it is in randomized experiments because biases particular to the circumstances of one study may not be present in other studies (Rosenbaum 2002). Our results will be strengthened if they can be replicated with different samples collected in different places at different times. Here we replicate our findings using a survey conducted in 1984.[30] It contains about 3000 respondents from several smaller East German cities (Halle, Schwerin, Dessau, Magdeburg, Bitterfeld, Greifswald) and also 950 respondents from the Dresden district.

We have attempted to make this replication as similar as possible to the analysis presented in the paper. For now, we exclude Greifswald respondents from the sample since many of them had no access to West German television (see below). The instrument is again coded 0 for respondents from the Dresden district and 1 for respondents from other districts. The treatment indicator is again based on a question that asks respondents how often they watch West German television. The six response categories are daily, two to five times per week, once per week, once or twice per month, less often, and never. The treatment indicator is coded 1 for respondents who

[29] It was well known that living conditions in East Berlin were better than in the rest of East Germany. Many of the outliers we see in Figure 2 are due to East Berlin.

[30] Zentralarchiv für Empirische Sozialforschung ZA 6129: Friedensstudie 1983.

fall in the first four response categories and 0 for all others.[31] The correlation between exposure to West German television and place of residence is again very high (.40).

We have chosen outcome variables that are as close as possible to the questions in the 1988/89 survey:

- *"I am proud of being a citizen of our socialist country."*
- *"It is personally important for me to help advance socialism."*
- *"Socialism can only succeed if workers and farmers have a firm grasp on political power under the leadership of the communist party."*

The four response categories are similar to the ones used in the 1988/89 survey (strongly disagree to strongly agree). We again use a limited and an extensive set of covariates.[32] As before, we discretize all variables to avoid functional form assumptions.

Results are displayed in Table 5.8, which shows treatment effect estimates for all three outcome variables and both covariate sets. Standard errors are again adjusted for clustering. To economize on space and given the similarity between LARF and 2SLS results, we only present 2SLS results here.

We find that across all models, West German television exposure has a positive effect on East Germans' support for the communist regime. Effect sizes are comparable to the results reported in Table 3 in the paper; all estimates are statistically significant at conventional levels. Even though the survey was conducted in different districts five years earlier and survey questions differ somewhat from the questions in the 1988/89 survey, we can fully replicate our results.

[31]Results are similar if respondents in the first 3 response categories are coded as treated or if only respondents in the last response category are coded as controls.

[32]The limited set contains age and gender (no information on parents' occupational classification is available). The extensive set adds educational attainment, occupational classification, marriage status, number of children, and employment status (no information on net income is available) to the limited set.

5.8.4 Replication II: Alternative control group

The 1984 survey allows for another type of replication. It contains 167 respondents from the town of Greifswald, which had limited access to West German over-the-air television broadcasts due to its location in the Northeastern corner of East Germany (cf. Figure 1 in the paper). If we can replicate our results with this group of control respondents, the possibility is ruled out that the results reported in the paper are driven by some unobserved confounder particular to the Dresden district.

We drop all respondents from the Dresden district for this replication; only Greifswald respondents are assigned to the control group. Note that this dramatically reduces the sample size for the controls from 950 to 167. Although the correlation is lower than before, West German television exposure and place of residence are still correlated (.15). We fit the same specifications as in the previous section.

Table 5.9 displays the estimates. Similar to the results in Table 5.8, we again see a positive effect of West German television exposure on support for the communist regime. The estimated treatment effects are substantively important for all outcome variables and both covariate sets.[33] Half of the estimates are significant at conventional levels despite the small number of control respondents. The fact that we again find a positive treatment effect — even though we are using respondents living in another East German city as control group — demonstrates the robustness of our results.

[33]The effect sizes are larger than the corresponding estimates in Table 5.8. However, note the larger standard errors in Table 5.9, which are due to the small number of Greifswald respondents. For none of the outcome variables can we reject the null hypothesis that the point estimates in Tables 5.8 and 5.9 are identical.

5.9 Tables for Chapter 5

Table 5.1: West German television exposure and place of residence

	Living in			
How often do you	Dresden district		other districts	
watch West German television?	Count	%	Count	%
daily	42	5	1820	65
multiple times each week	69	9	732	26
once a week	17	2	47	2
less than once a week	108	14	101	4
never	498	64	46	2
Missing	50	6	33	1
Total	784	100	2779	100

Table 5.2: Outcome variables

	fully disagree	largely disagree	largely agree	fully agree	Missing	Total
Convinced of Leninist/	399	850	1697	571	47	3564
Marxist worldview	11%	24%	48%	16%	1%	100%
Feel closely attached	154	600	1963	809	38	3564
to East Germany	4%	17%	55%	23%	1%	100%
Political power is exercised in	301	874	1969	373	47	3564
ways consistent with my views	8%	25%	55%	10%	1%	100%

Table 5.3: Effect of West German television exposure on regime support

Estimator	Diff.	LATE	2SLS	LARF	2SLS	LARF
Covariate set	—	—	Limited	Limited	Extensive	Extensive
Convinced of Leninist/Marxist worldview						
West German tv	-0.079	0.147	0.205	0.204	0.198	0.204
Standard error	(0.053)	(0.083)	(0.084)	(0.084)	(0.087)	(0.108)
Feel closely attached to East Germany						
West German tv	-0.013	0.217	0.258	0.255	0.256	0.251
Standard error	(0.044)	(0.067)	(0.072)	(0.075)	(0.073)	(0.090)
Political power exercised in ways consistent with my views						
West German tv	-0.014	0.158	0.193	0.191	0.186	0.185
Standard error	(0.047)	(0.078)	(0.082)	(0.083)	(0.081)	(0.106)

Note: $N = 3441$ (3426) for models without (with) covariates. Treatment effect estimates for each estimator are shown with robust cluster-adjusted standard errors in parentheses. Diff. denotes the difference in means between exposed and unexposed respondents. LATE is the unconditional average treatment effect for compliers. 2SLS is the two-stage least squares estimate. LARF is the estimate from a local average response function. The limited covariate set includes age, gender, and father's and mother's occupational classification. The extensive covariate set adds respondent's marriage status, living situation, number of children, highest educational attainment, occupational classification, net monthly income, and employment status to the limited set. Response categories for the outcome variables are coded as fully disagree = 1, largely disagree = 2, largely agree = 3, and fully agree = 4.

Table 5.4: Deterrence and escapism

Do you feel attached to West Germany? (mean = 2.28; sd = 1.08)		
Covariate set	Limited	Extensive
West German tv	-0.193	-0.172
Standard error	(0.111)	(0.113)
N	1141	1141

How satisfied are you with the available options for recreational activities? (mean = 3.05; sd = 1.10)		
Covariate set	Limited	Extensive
West German tv	0.395	0.405
Standard error	(0.117)	(0.110)
N	1148	1148

Note: Results are 2SLS estimates with cluster-adjusted standard errors in parentheses. The limited covariate set includes age, gender, and father's and mother's occupational classification. The extensive covariate set adds respondent's marriage status, living situation, number of children, highest educational attainment, occupational classification, net monthly income, and employment status to the limited set. The 5-point response categories range from not at all = 1 to fully = 5.

Table 5.5: Exit visa application rates in 1988

District	Rate
Dresden	16.4
East Berlin	11.8
Karl-Marx-Stadt	9.8
Leipzig	7.9
Erfurt	5.0
Cottbus	3.5
Magdeburg	2.8
Schwerin	2.4
mean	7.45

Note: Number of exit visa applications per 1,000 residents in 1988. Source: BStU MfS-ZKG 10734.

Table 5.6: Effect of West German television exposure on regime support: First differences from ordered probit LARF estimator

Predicted shift in $Pr(Y = j)$	j = 1 fully disagree	j = 2 largely disagree	j = 3 largely agree	j = 4 fully agree
Convinced of Leninist/Marxist worldview				
West German tv	−0.038	−0.050	0.022	0.065
Upper CI bound	(−0.000)	(−0.004)	(0.155)	(0.145)
Lower CI bound	(−0.069)	(−0.267)	(0.002)	(0.008)
Feel closely attached to East Germany				
West German tv	−0.026	−0.073	−0.028	0.127
Upper CI bound	(−0.000)	(−0.013)	(0.115)	(0.261)
Lower CI bound	(−0.125)	(−0.303)	(−0.080)	(0.041)
Political power exercised in ways consistent with my views				
West German tv	−0.036	−0.064	0.044	0.056
Upper CI bound	(−0.000)	(−0.004)	(0.180)	(0.117)
Lower CI bound	(−0.063)	(−0.264)	(0.003)	(0.004)

Note: $N = 3426$ for all models. Treatment effect estimates with upper and lower bounds for robust cluster-adjusted .95 confidence intervals in parentheses. Estimates are average treatment effects for compliers in first differences, i.e., the average changes in the predicted $Pr(Y = j)$ for compliers under treatment and control. All models include the limited covariate set.

Table 5.7: Effect of West German television exposure on regime support (subsamples)

	Outcome variable			
Treated district	Y1	Y2	Y3	N
Schwerin	0.285	0.490	0.350	976
	(0.130)	(0.109)	(0.123)	
Magdeburg	0.322	0.399	0.223	1333
	(0.074)	(0.066)	(0.070)	
East Berlin	0.133	0.099	-0.006	1203
	(0.084)	(0.074)	(0.074)	
Leipzig	0.338	0.498	0.435	1061
	(0.100)	(0.087)	(0.086)	
Karl-Marx-Stadt	0.147	0.219	0.212	1417
	(0.075)	(0.064)	(0.065)	
Erfurt	0.137	0.151	0.123	1433
	(0.073)	(0.065)	(0.067)	

Note: Y1 = *Convinced of Leninist/Marxist worldview.* Y2 = *Feel closely attached to East Germany.* Y3 = *Political power exercised in ways consistent with my views.* Local average treatment effect estimates for compliers from linear LARF estimator. Robust cluster-adjusted standard errors are in parentheses. All models include the limited covariate set. Each row presents estimates for a different subgroup, i.e., estimates for a sample with Dresden and one other district as the treatment group.

Table 5.8: Effect of West German television exposure on regime support: Replication with 1984 survey

I am proud of being a citizen of our socialist country		
Covariate set	Limited	Extensive
West German tv	0.467	0.465
Standard error	(0.113)	(0.106)
N	2901	2901
It is personally important for me to help advance socialism		
Covariate set	Limited	Extensive
West German tv	0.366	0.390
Standard error	(0.128)	(0.128)
N	2897	2897
Socialism can only succeed if workers and farmers have a firm grasp on political power under the leadership of the communist party		
Covariate set	Limited	Extensive
West German tv	0.137	0.146
Standard error	(0.073)	(0.073)
N	2897	2897

Note: Results are 2SLS estimates with robust cluster-adjusted standard errors in parentheses. The limited covariate set includes age and gender. The extensive covariate set adds educational attainment, occupational classification, marriage status, number of children, and employment status to the limited set. Response categories are coded as fully disagree = 1, largely disagree = 2, largely agree = 3, and fully agree = 4. The control group consists of 950 respondents living in the Dresden district.

Table 5.9: Effect of West German television exposure on regime support: Replication with 1984 survey and Greifswald respondents as control group

I am proud of being a citizen of our socialist country		
Covariate set	Limited	Extensive
West German tv	1.770	1.218
Standard error	(0.843)	(0.864)
N	2155	2155

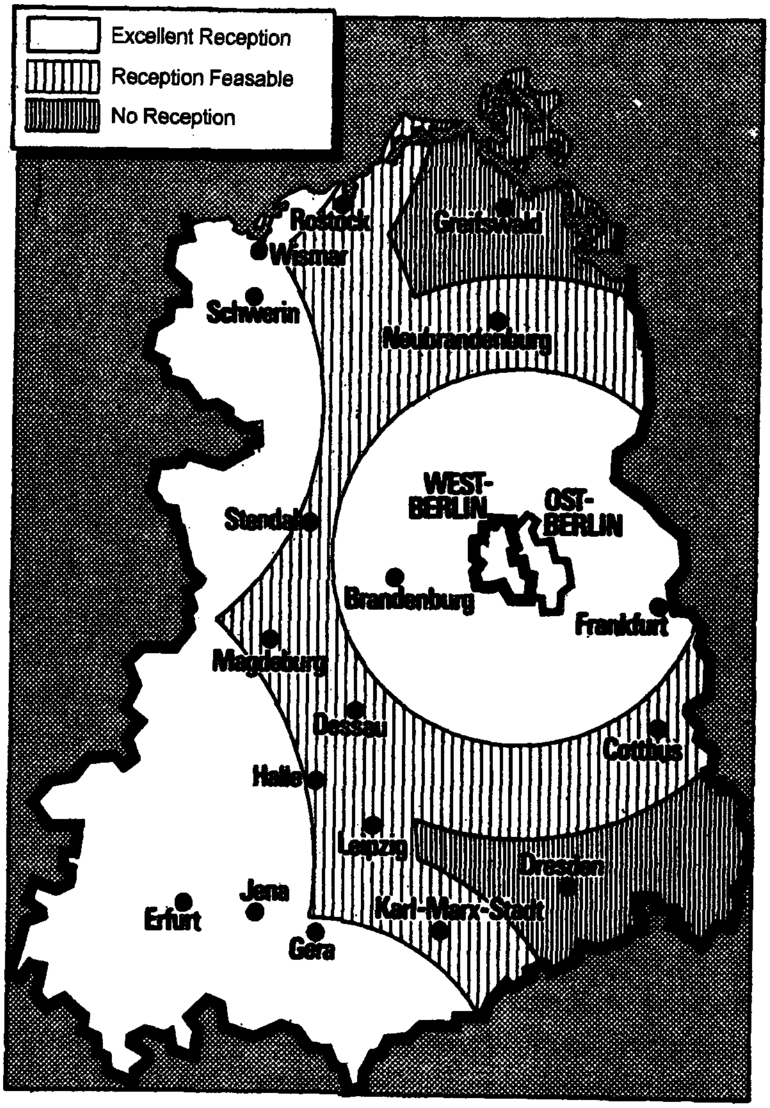
It is personally important for me to help advance socialism		
Covariate set	Limited	Extensive
West German tv	1.598	1.920
Standard error	(1.564)	(1.276)
N	2154	2154

Socialism can only succeed if workers and farmers have a firm grasp on political power under the leadership of the communist party		
Covariate set	Limited	Extensive
West German tv	0.645	0.856
Standard error	(0.380)	(0.382)
N	2156	2156

Note: Results are 2SLS estimates with robust cluster-adjusted standard errors in parentheses. The limited covariate set includes age and gender. The extensive covariate set adds educational attainment, occupational classification, marriage status, number of children, and employment status to the limited set. Response categories are coded as fully disagree = 1, largely disagree = 2, largely agree = 3, and fully agree = 4. The control group consists of 167 respondents living in Greifswald.

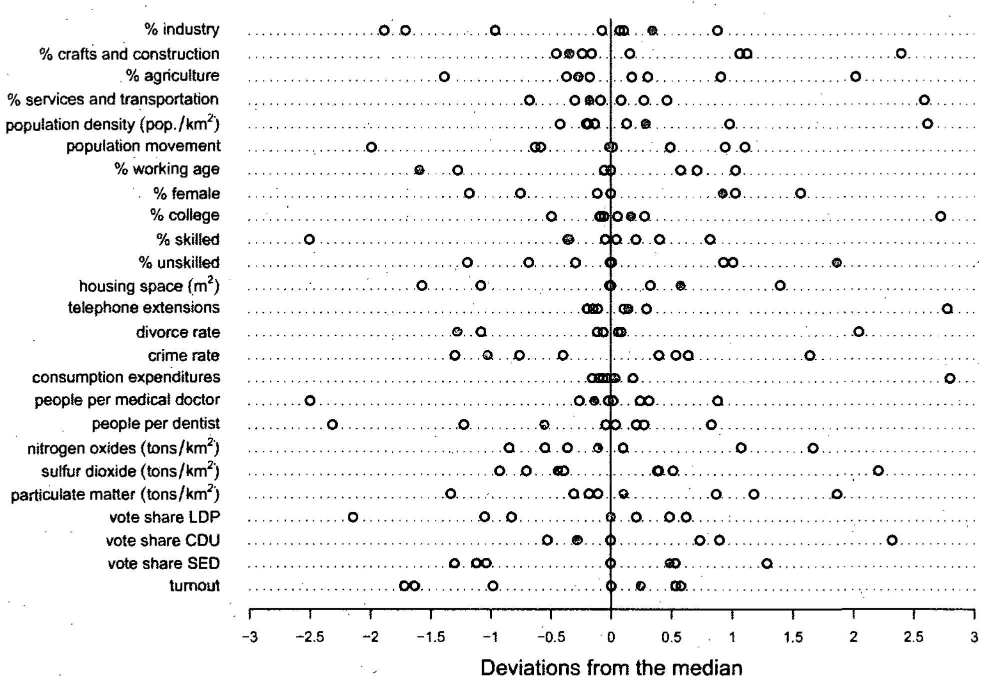
5.10 Figures for Chapter 5

Figure 5.1: Over-the-air signal strength of West German television broadcasts



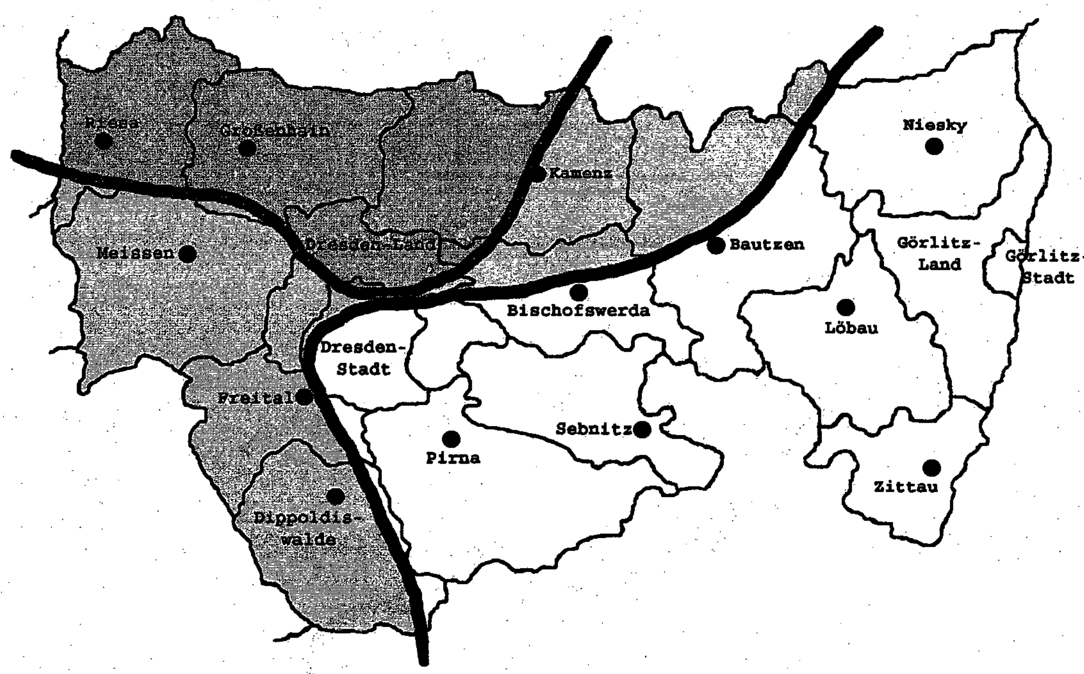
Source: Adapted from *Die Welt*, October 27, 1980.

Figure 5.2: District characteristics



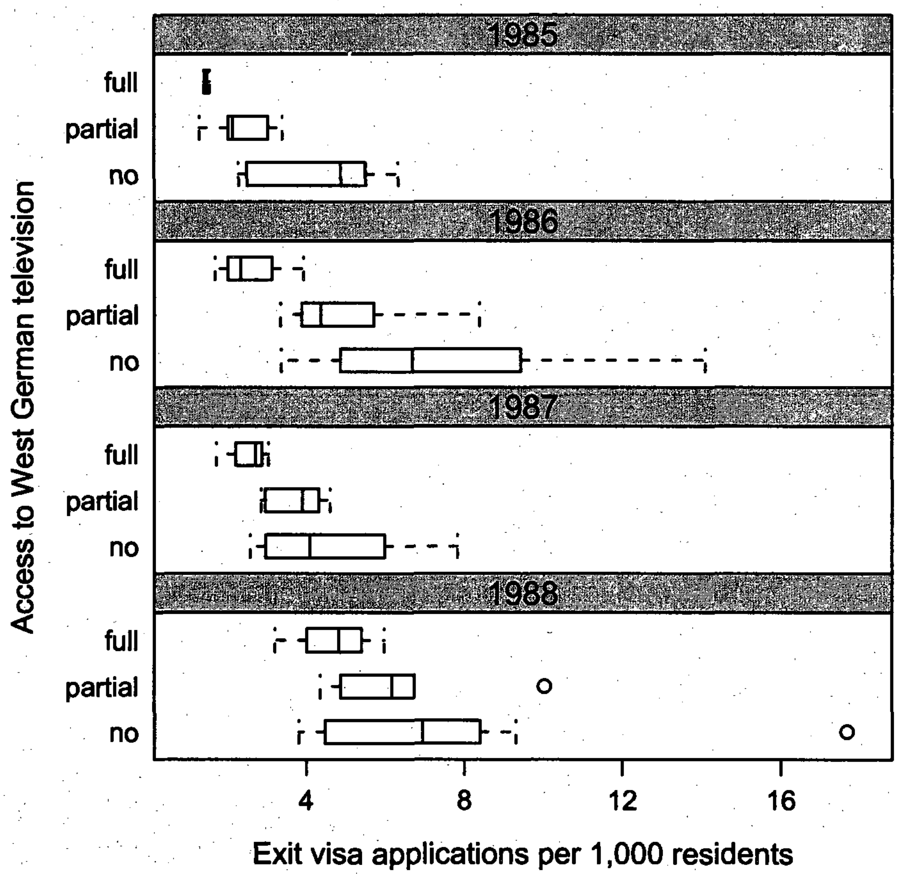
Note: The graph shows balance in district characteristics for the 8 districts in our sample. The x-axis shows deviations from the medians; the y-axis lists the district characteristics. Red solid circles display values for the Dresden district; empty circles display values for the other districts. Note that for several district characteristics only seven values are shown as data for East Berlin are not always available. Data are taken from Staatliche Zentralverwaltung für Statistik 1989 and Grundmann 1997.

Figure 5.3: Over-the-air signal strength of West German television within the Dresden district



The map shows spatial patterns in access to West German television across the 17 counties in the Dresden district. White areas had no access, areas shaded in light grey had partial access, and areas shaded in dark grey had full access to West German television. Source: Adapted from original map in BArch SAPMO DY 30/J IV2/2/2317: 75.

Figure 5.4: Exit visa application rates and access to West German television for counties in the Dresden district



Each of the four panels shows three box-and-whisker plots for counties with varying levels of access to West German television. Source: BStU MfS BV Dresden Stellvert. OP 21; BStU MfS BV Dresden BKG-8028/8029.

Chapter 6

Economic Concerns and Attitudes Towards Immigration

6.1 Introduction

Why do people oppose or favor immigration? Recent scholarly work examining survey data on individual attitudes towards immigration has generated inconsistent findings and no clear consensus view. Some studies suggest that opposition to immigration is primarily driven by non-economic concerns associated with cultural and ethnic differences between native and immigrant populations (Dustmann & Preston 2007, Dustmann & Preston 2006, Espenshade & Hempstead 1996, Chandler & Tsai 2001, Citrin, Green, Muste & Wong 1997, Bauer, Lofstrom & Zimmerman 2000, Burns & Gimpel 2000, Fetzer 2000, Gang, Rivera-Batiz & Yun 2002). Others argue that economic concerns lie at the heart of anti-immigrant sentiment and that individual attitudes towards immigration are profoundly shaped by fears about labor market competition (Scheve & Slaughter 2001, Kessler 2001, Mayda 2006) and the fiscal burden on public services (Facchini & Mayda 2007, Hanson 2005, Hanson, Scheve & Slaughter 2007). Borjas (1999) identifies these as the two economic issues that have dominated the debate over immigration policy in the United States. But there is no agreement among scholars about the relative impact of these types of concerns or how they compare in importance with non-economic considerations that also motivate anti-immigrant sentiment. These issues are critical for understanding public opposition to immigration and the growth of extremist, often violent, anti-immigrant political movements.

One reason there is no consensus on why people support or oppose immigration is that the data on individual attitudes is ill-suited to testing the theoretical relationships at issue. Studies examining economic concerns about immigration typically begin with a general equilibrium model and derive predictions about how native citizens who own different types of productive factors, and who have different levels of income, will differ in their views regarding highly skilled and low skilled immigration (Scheve & Slaughter 2001, Mayda 2006, Facchini & Mayda 2007, Hanson et al. 2007). However, due to data constraints, none of these studies have been able to test these specific predictions directly. They rely instead upon indirect tests that leave the interpretation of the results wide open. In particular, no study to date has been able to distinguish between attitudes towards highly skilled immigrants and attitudes towards low skilled immigrants, even though this distinction is a critical feature of the theoretical story about how economic concerns affect attitude formation and policy preferences with respect to immigration.

We conducted a unique survey experiment that, for the first time, explicitly and separately examines individuals' attitudes towards highly skilled and low skilled immigrants. In a nationwide U.S. survey, we randomly assigned respondents to answer questions about immigrants with different skill levels, thereby obtaining an unbiased comparison between the distributions of attitudes towards highly skilled and low skilled immigrants. This comparison, and how it varies with respondent characteristics, allows us to directly test the predictions from the theoretical models about how economic concerns affect attitudes towards immigration.

The experiment yields results that challenge existing models and the conclusions reached by previous studies of attitudes towards immigration. The prominent *labor* *market competition model* predicts that natives will be most opposed to immigrants who have similar skill levels to their own. This is rejected by the data. We find that *both* highly skilled and low skilled respondents strongly prefer highly skilled immigrants over low skilled immigrants, and this preference is not decreasing in respondents' skill levels. Support for *both* highly skilled and low skilled immigration is strongly increasing in respondents' skill levels. In addition, these relationships are similar for the sub-samples of respondents that are currently in or currently out of the labor force. The results suggest that, among natives generally, labor market competition is not a significant motivator of anti-immigrant sentiment.

The *fiscal burden model* anticipates that rich (high income) natives oppose low skilled immigration and favor highly skilled immigration more than do poor (low income) natives, and that this difference should be more pronounced in states with greater fiscal exposure in terms of immigrant access to public services. We find instead that rich and poor natives are *equally* opposed to low skilled immigration and supportive of highly skilled immigration most of the time, although poorer natives are actually significantly more opposed to low skilled immigration than richer counterparts in states with high fiscal exposure. In addition, rich natives are actually less opposed to low skilled immigration in states with more fiscal exposure than they are elsewhere. These results are inconsistent with claims that rich natives are opposed to low skilled immigrants because they anticipate a heavier tax burden associated with the provision of public services. Finally, we find evidence that poor natives are more opposed to low skilled immigration in states with greater fiscal exposure than they are elsewhere. Taken together, these results suggest that concerns about access to or overcrowding of public services contributes to anti-immigrant attitudes among poorer native citizens.

6.2 Economically Motivated Attitudes Toward Immigration

Although immigration may impact the native economy in many ways, recent research has emphasized two critical economic concerns that could generate anti-immigrant sentiment among native citizens: concerns about labor market competition and fears about the fiscal burden on public services. General equilibrium models of the native economy generate a variety of predictions about how natives with particular skill and income characteristics should be affected by inflows of immigrants.

6.2.1 Labor Market Competition

Analysis of the income effects of immigration typically begins with a closed-economy general equilibrium "factor-proportions" (FP) analysis (Borjas, Freeman & Katz 1996, Borjas 1997, Borjas 1999). The FP model derives the distributional effects in the native economy from the impact that immigration has on the relative supplies of factors of production. If immigrants have low skill endowments compared with natives, immigration will raise the supply of low skilled labor relative to other factors (including highly skilled labor). These changes in relative factor supplies translate into changes in real factor returns: wages of native low skilled workers will fall as new (low skilled) immigrants price themselves into employment; and, as more low skilled labor is applied to fixed amounts of the other factors, the real wages of highly skilled workers will rise. The reverse effects are expected in the case of inflows of highly skilled immigrants, which will drive up the real wages of low skilled natives while reducing real returns for highly skilled natives. Depending on what one assumes about wage flexibility, the impact of competition with similarly-skilled immigrants may also be manifest in higher rates of unemployment among natives.[1] The FP model generates a clear prediction about attitudes towards immigration: natives should oppose

[1] Alternative models also allow for geographic concentration of wage and employment effects. See Card (1990) or Borjas (1999).

immigrants with similar skill levels but favor immigrants with different skill levels. (See appendix A for formal derivations of these relationships).

Empirical studies have found mixed results when testing this model (Harwood 1986, Gang et al. 2002, Fetzer 2000, Burns & Gimpel 2000, Citrin et al. 1997, Dustmann & Preston 2006), although two prominent articles have recently reported strong supporting evidence. Drawing upon data from the National Election Studies (NES) surveys in the United States in the 1990s, Scheve & Slaughter (2001) find a strong positive correlation between respondents' skill levels, as measured by years of education, and stated support for immigration. This correlation is interpreted as evidence that low skilled (less educated) natives fear being forced to compete for jobs with low skilled immigrants. In a similar study Mayda (2006) examined cross-national survey data from the 1995 National Identity Module of the International Social Survey Programme (ISSP), as well as data collected between 1995 and 1997 by the World Value Survey (WVS), and finds that the probability of voicing pro-immigration opinions is positively associated with the skill levels of survey respondents (measured by years of education). Again, this correlation is presented as confirmation that concerns about labor market competition are a powerful motivator of attitudes towards immigrants.

There are several reasons to be wary of these reported findings. First, it is unclear whether respondents can plausibly observe and correctly attribute the income effects of immigration that are anticipated in the FP model. A growing set of empirical studies has examined the effect of immigration on native wages and unemployment, but the evidence remains hotly debated.[2] Some studies claim large, adverse wage and employment effects of immigration on less educated workers (Borjas et al. 1996, Borjas 1997, Borjas 1999, Borjas 2003, Borjas 2005), while others conclude that the immigration effects are at most very small, and possibly insignificant (Card 1990, Card 2001, Card 2007, Lewis 2005). In a recent study Ottaviano & Peri (2008)

[2]For general reviews about the impact of immigration on wages and employment see for example Friedberg & Hunt (1995), Bhagwati (2002), Card (2005), Borjas (1999), and Longhi, Nijkamp & Poot (2005). In a recent study Borjas (2003) summarizes the evidence observing that "the measured impact of immigration on the wage of native workers fluctuates widely from study to study (and sometimes even within the same study) but seems to cluster around zero."

find a net positive long-term effect of immigration on average wages of natives. The inconclusiveness of the empirical research on the labor market effects of immigration suggests the need for caution when applying the simple FP model to make predictions about attitude formation and interpreting the evidence on attitudes.

Second, in line with the mixed empirical evidence on the impact of immigration, many scholars have pointed out that when we move away from the FP analysis and consider more sophisticated economic models, it becomes very difficult to make clear predictions about the equilibrium effects of immigration on wages and employment opportunities among native workers (see Friedberg & Hunt (1995); Scheve & Slaughter (2001), 135-7). In an open-economy Heckscher-Ohlin model, trade can offset the impact of immigration as the mix of output of tradable goods changes in line with changes in factor supplies. Assuming that the local economy is not large relative to the rest of the world and/or inflows of immigrants are small relative to the local labor supply, local wages will not be affected - the "factor price insensitivity" result (Leamer & Levinsohn 1995). In an amended open-economy model in which skills of workers are highly specific to particular industries (Jones 1971, Grossman & Helpman 1994), real income effects are sensitive to the inclusion of non-traded goods. Immigration can lead to a reduction in the price of non-traded goods (by raising the output of such goods more rapidly than it raises aggregate demand for them) and so it becomes unclear whether native workers with skills similar to those of immigrants will be worse off in real terms (this will depend in part on their consumption tastes). In alternative types of open-economy models which allow for economies of scale in production in the industries employing immigrants, inflows of new workers can be shown to generate higher real wages for native workers with similar skills (Brezis & Krugman 1993). There is, in short, a great deal of theoretical ambiguity about the labor market effects of immigration and the related concerns we should expect to observe among native citizens.

Third, a variety of alternative explanations can account for the positive correlation between education and pro-immigration attitudes. Several studies have shown that more educated respondents tend to exhibit higher levels of ethnic and racial tolerance, stronger preferences for cultural diversity, and more economic knowledge, all of which can lead them to favor immigration more than their less educated counterparts (Hainmueller & Hiscox 2007, Gang et al. 2002, Fetzer 2000, Chandler & Tsai 2001, Dustmann & Preston 2007, Citrin et al. 1997). Existing tests are not equipped to discriminate between these claims and the argument that the association between education and views about immigrants is due to concerns about labor market competition.[3]

Fourth and finally, all the above-mentioned tests that have examined attitudes towards immigration and tried to link them to concerns about labor market competition have relied upon data on responses to survey questions that ask individuals about their attitudes toward immigration *in general* and do not differentiate between highly skilled or low skilled immigrants.[4] This is highly problematic because the key prediction of the simple FP model is that natives should oppose immigrants with similar skill levels to their own but support immigrants with different skill levels. Previous tests rely on an assumption that respondents have *low* skilled immigrants in mind when answering questions about immigration in general. This assumption is questionable given that respondents are likely to have systematically varying information and perceptions about the skill attributes of immigrants. More educated respondents may be better informed about current immigration flows, for instance, and are likely to recognize the considerable share of inflows accounted for by skilled foreigners entering many Western nations (often because immigration policies are aimed explicitly at selecting immigrants based on their skill levels). It is well known

[3]The same problem applies to studies that test these models with respect to attitudes towards international trade. See Hainmueller & Hiscox (2006).

[4]Scheve & Slaughter (2001) used responses to the NES immigration question: "Do you think the number of immigrants from foreign countries who are permitted to come to the United States to live should be increased a little, increased a lot, decreased a little, decreased a lot, or left the same as it is now?" Mayda (2006) examined answers to the ISSP question: "Do you think the number of immigrants to (respondents country) nowadays should be: (a) reduced a lot, (b) reduced a little, (c) remain the same as it is, (d) increased a little, or (e) increased a lot." The WVS asked the following question: "How about people from other countries coming here to work. Which one of the following do you think the government should do (a) Let anyone come who wants to (b) Let people come as long as there are jobs available (c) Place strict limits on the number of foreigners who can come here (d) Prohibit people coming here from other countries? (e) Don't know."

that such varying perceptions can lead to biased estimates in survey research (King, Murray, Salomon & Tandon 2004, Bertrand & Mullainathan 2001). And of course, employing this questionable assumption still does not allow one to examine whether the skill levels of natives affects their attitudes towards *highly* skilled immigrants in the expected way. A complete and direct test would ask respondents about their attitudes towards low skilled immigrants and highly skilled immigrants specifically and separately.

The only previous study that comes close to such a test actually reports results at odds with the recent claims that labor market concerns are powerful shapers of attitudes. Hainmueller & Hiscox (2007) investigate survey data for 22 European countries from the European Social Survey (ESS), in which respondents were asked about their attitudes towards immigration from "richer" and "poorer" countries, a difference plausibly associated with the expected average skill levels of immigrants. They find that in all 22 countries people with higher education levels (and/or higher levels of occupational skills) are more likely to favor immigration regardless of where the immigrants come from and their likely skill attributes. In addition, the positive link between education and support for (all types of) immigration is almost identical among those in the labor force and those not in the labor force. Taken together, the existing theory and evidence on whether concerns about labor market competition are a strong motivator of anti-immigrant sentiment remain ambiguous. At the very least, more complete and direct empirical tests are necessary.

6.2.2 The Fiscal Burden of Public Services

The second critical economic concern associated with immigration involves the immigrants' use of public services (including public education and health services, and various types of welfare assistance, as well as basic services like police and fire protection, roads, parks, and amenities) and their contribution to tax revenues. The standard approach to the analysis is to incorporate a simple model of public finance into the FP analysis of immigration (Facchini & Mayda 2007, Hanson 2005, Hanson et al. 2007). This approach allows that immigration can affect not only the pre-tax incomes of native individuals, it can separately affect after-tax incomes via taxes and transfers. The predictions depend on two key assumptions about (1) the net contribution of low and highly skilled immigrants to the tax coffers and (2) the institutional mechanism in place to adjust taxes and transfers in response to fiscal imbalances. It is assumed that low skilled immigrants impose a substantial net burden on public finance, while highly skilled immigrants are net contributors in terms of taxes. There are two plausible institutional mechanisms that have been considered, assuming the government must balance its budget: a change in tax rates or a change in per capita transfers (see Facchini & Mayda (2007)).[5] In the most commonly studied scenario, assuming the government adjusts tax rates while keeping per capita transfers constant, the prediction is that rich (high income) natives should prefer highly skilled over low skilled immigrants more than do poor (low income) natives, since the skill levels of immigrants determine their fiscal impact and progressivity in taxation implies that the rich benefit (lose) more from any associated reduction (increase) in taxes. In the alternative scenario, assuming the government adjusts per capita transfers but holds tax rates constant, the prediction is the opposite: poor natives prefer highly skilled over low skilled immigrants more than rich natives, since low skilled immigrants tend to crowd out poor natives in terms of access to public services and erode their welfare benefits while rich natives are unaffected. (See appendix A for formal derivations of these relationships).

Two recent empirical studies have examined these claims. Hanson et al. (2007) use NES survey data to compare individual attitudes towards immigration in different U.S. states and find that rich individuals are less likely to support immigration in states that are highly exposed to fiscal costs as a result of immigration (i.e., states with generous public services and high rates of immigrant settlement) than in states with lower exposure. This finding is interpreted as confirmation that, as expected in the scenario in which the government adjusts taxes to meet new spending obligations, rich natives fear being burdened with higher taxes as a consequence of low skilled

[5]Borrowing would be a third option, but as there are constitutional constraints on borrowing by state governments in the United States, and the underlying model is static, standard analyses do not consider this possibility (Hanson et al. 2007, Facchini & Mayda 2007).

immigrants drawing on public services and draining government coffers. Facchini & Mayda (2007) examine the cross-national survey data from the ISSP and find that respondent income is negatively correlated with support for immigration in countries where low skilled immigrants are a larger share of total immigration inflows. This finding is also regarded as evidence that fears about higher taxes among rich natives, linked to use of public services by low skilled immigrants, leads to anti-immigrant sentiments.

Again, there are reasons to treat these findings with considerable caution. While there is some evidence that immigrants rely more on welfare programs than do native citizens (Borjas 1999, Zimmerman & Tumlin 1999, Fix & Passel 2002, Hanson 2005), as immigrant households tend to be larger and poorer than native households, there is more disagreement over the extent to which immigrant inflows increase net tax burdens on natives (Fix, Passel & Enchautegui 1994, Smith & Edmonston 1997). A U.S. study conducted by the National Research Council (NRC) reported that the average immigrant to the United States could be expected to impose a tax burden on natives in the short term, but would be a net contributor to tax coffers in the long term, to the tune of $80,000 (see Smith & Edmonston (1997)).[6] Estimating the long term fiscal consequences of immigration in a dynamic model of public finance is very difficult, of course, and requires taking into account fiscal contributions made by successive generations of immigrant and native families. For countries with aging workforces, in particular, the long term public finance gains from importing young workers likely outweigh the costs (Krugman & Obstfeld 2000). Perhaps short term fiscal effects dominate longer term effects in shaping attitudes among native citizens, but the evidence is complicated enough to suggest caution when claiming that fears about the tax effects of immigration are a strong motivation for anti-immigrant sentiments.

Quite separately, the finding that tax considerations among natives play a strong

[6]The study reports findings in 1996 dollars. The NRC study did report that tax affects vary depending on the skill levels of immigrants: immigrants with an education beyond high school contribute an average of $105,000 to U.S. tax coffers over their lifetime, while the least educated immigrants create a net deficit of $89,000 per person (Smith & Edmonston 1997).

role, and actually trump concerns about cuts in per capita welfare benefits, seems especially surprising in the United States. Evidence on recent fiscal experiences of U.S. states seems inconsistent with this claim. While states gained broad discretion over welfare policies following the welfare reform of 1996, they have not systematically raised taxes in recent years even though immigration has increased. In fact, as shown in left panel of Figure 6.1, looking across the states there exists, if anything, a negative correlation between changes in state income tax rates and levels of immigration. States that experienced higher increases in their foreign born populations between 1990 and 2004 had smaller increases (or larger cuts) in the average marginal tax rates than states with smaller immigrant inflows over the same time period.[7] It seems unlikely, then, that U.S. survey respondents could be drawing on personal experience to attribute tax hikes to immigration.

On the other hand, a recent study that looks at the link between immigration and U.S. state welfare expenditures has found stronger support for the so called "erosion hypothesis." Hero & Preuhs (2007) examine data on welfare spending for all U.S. states in 1998 and find that states with larger noncitizen populations tend to provide smaller cash benefits in their welfare programs, and this effect is larger the more accessible the welfare programs are to immigrants. In the right panel of Figure 6.1 we plot changes in state public welfare expenditures per capita against changes in the immigrant population. There is a negative correlation between the two. While all states have expanded per capita welfare expenditures over time, the increases have been smaller in states that experienced larger increases in the share of immigrants in their population.[8] This pattern, taken together with the evidence on state taxes

[7]For both tax rates and the percent foreign born population, changes are computed as the level in 2004 minus the level in 1990. Tax rates are average marginal state tax rates on wages taken from the NBER state tax database (Feenberg & Coutts 1993) available at `http://www.nber.org/~taxsim/state-marginal/`. Income taxes are dollar weighted average marginal income tax rates as calculated by the NBER TAXSIM model from micro data for a sample of U.S. taxpayers. The results are very similar if tax rates on other sources of income are used (i.e., taxes on interest received, dividends, pensions, or property tax, etc.). Data on the percent foreign born is taken from the U.S. Census 1990 and the American Community Survey 2004.

[8]Public welfare expenditures are taken from the U.S. Census of Governments (see the following section for more details on the welfare spending measures).

discussed above, suggests that fears about the erosion of welfare benefits as a result of immigration may actually be more relevant and plausible than worries about tax hikes.

Finally, it should be noted that the survey-based tests summarized above are indirect and incomplete. Like the studies that examine concerns about labor market competition, existing tests of the fiscal burden model rely upon data on responses to NES and ISSP survey questions that ask individuals about their attitudes toward immigration *in general*, not about their attitudes toward highly skilled or low skilled immigrants specifically. They rest on the problematic assumption that all respondents actually have low skilled immigrants in mind when answering these survey questions about immigration. And employing this assumption still only allows a partial test of the theory: it does not allow one to test whether the incomes of natives affects their attitudes towards *highly* skilled immigrants in the expected way.

In sum the existing research examining whether attitudes towards immigrants are strongly shaped by concerns about labor market competition and fears about the fiscal burden on public services does not provide convincing conclusions. Most importantly, as a result of data constraints, these studies have not been able to provide direct tests of the relevant theoretical propositions. No study to date has been able to distinguish between attitudes towards highly skilled immigrants and attitudes towards low skilled immigrants, even though this distinction is a critical feature of the theoretical story. Below we describe a survey experiment aimed at addressing this shortcoming.

6.3 The Survey Experiment

6.3.1 Design

Our experiment was embedded in the Cognitive Styles Survey (CSS), a survey instrument designed to study opinions regarding trade and immigration. The CSS was administered by the research firm Knowledge Networks (KN) and fielded as an internet based questionnaire to a randomly selected sample of 2,285 panel members between December 2007 and January 2008. KN recruits members for its research panel based on probability sampling that covers both the online and offline U.S. populations aged 18 years and older.[9] The sampling procedure for the CSS thus constitutes a two-stage equal probability design. The completion rate for the CSS was about 70 percent, yielding 1,600 completed interviews. The final respondent data was adjusted for the common sources of survey error (non-response, coverage error, etc.) using post-stratification weights.[10] The rate of item non-response was very low, below 1 percent for the questions we use in the analysis below.

For the core experiment, we randomly allocated respondents to two groups of equal size and presented each group with one of two versions of the survey question about immigration:

Version 1: Do you agree or disagree that the US should allow more **highly skilled** immigrants from other countries to come and live here? (emphasis added)

Version 2: Do you agree or disagree that the US should allow more **low skilled** immigrants from other countries to come and live here? (emphasis added)

Answer options (both versions):

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
1	2	3	4	5

The two question versions differed only in that they described the immigrants' skill level as either "*highly skilled*" or "*low skilled.*"[11] Accordingly, for half the re-

[9]Panel members are randomly selected using random digit dialing (RDD) sampling techniques on the sample frame consisting of the entire U.S. residential telephone population (both listed and unlisted phone numbers). Households are provided with access to the Internet and hardware if needed. Unselected volunteers are not allowed to join the KN panel. A detailed report about the KN recruitment methodology and the survey administration is available from the authors upon request.

[10]Poststratification also serves the purpose of improving precision of the estimates. Weights are raked to adjust to the demographic and geographic distributions from the March Supplement of the 2007 Current Population Survey (CPS).

[11]Notice that we stratified the random assignment by four education levels (described below) so that an equal number of respondents within each education level received the two different version

spondents, referred to as the treatment group, we measured preferences over highly skilled immigration while for the other half, referred to as the control group, we measured preferences over low skilled immigration. Randomization ensured that the two groups of respondents are (in expectation) identical in all other observed and unobserved characteristics that may confound a comparison across groups.[12]

The general distribution of preferences over both highly skilled and low skilled immigrants is displayed in Figure 6.2. For both types of immigration the barplots show the fraction of respondents answering each of the five answer categories; the superimposed whiskers decode the upper .95 confidence interval derived from the design based variance estimator. Two features stand out in this graph. First, in line with previous studies, our survey once again confirms the profound divide among the American public in opinions on immigration. Pooling over both types of immigration, about 50 percent of the respondents oppose an increase in immigration, while about 25 percent favor it. Second and more importantly, our findings for the first time document the fact that preferences over immigration vary rather dramatically depending on the immigrants' skill levels. While more than 60 percent of the respondents (in the control group) state that they strongly disagree or somewhat disagree with an increase in low skilled immigration, only 40 percent of the respondents (in the treatment group) are opposed to an increase in high skilled immigration.[13] Due to the randomization we know that this statistically significant difference between the two distributions is entirely driven by the perceived differences in the skill attributes of the immigrants.

---

of the question.

[12]We conducted extensive balance checks by comparing the distributions of all our covariates in both groups. All tests confirmed that (as expected given the large sample size) randomization balanced the distributions evenly. Results are available upon request.

[13]In the pre-implementation pilot testing, we created a third, "vanilla" version of the question that referred simply to "'immigrants"', without mentioning skill levels, and we randomly assigned respondents into a third group who answered this question. Opposition to immigration among this group was lower than opposition to low skilled immigration (in the pilot control group), and higher than opposition to highly skilled immigration (in the pilot treatment group). Since the results fell in the middle when no skill levels were specified, we focused on just the two contrasting versions of the question when we implemented the survey experiment.

In an additional experiment we replicated all our tests based on within-group variation by using a cross-over design. For this follow-up test, we contacted a random subset of the respondents two weeks after they had completed the main survey. Half of these respondents we randomly selected to receive the alternate version of the question they had received in the original survey two weeks prior. This approach allowed us to explore individual responses to both questions while minimizing the danger of "consistency bias."[14] The results from the analysis of this follow-up experiment, which strongly confirm the results from the main experiment reported below, are described in Appendix B.

6.4 Empirical Test I: The Labor Market Competition Model

6.4.1 Skill Levels of Natives

If concerns about labor market competition are important in shaping attitudes towards immigration we expect, in line with the FP model of attitude formation, that natives should oppose immigrants with similar skill levels but favor immigrants with different skill levels. That is, we should expect that the skill levels of our survey respondents should have a large and *positive* relationship with support for *low skilled* immigrants and a large and *negative* effect on support for *highly skilled* immigrants.

In order to conduct an explicit test of this argument we follow previous studies and employ educational attainment as our measure of respondent skill levels (Mayda & Rodrik 2005, Facchini & Mayda 2007, Scheve & Slaughter 2001, Hanson, Scheve & Slaughter 2008, Hanson et al. 2007). This measure, which we label EDUCATION, is a categorical indicator of the highest level of education attained by the respondent. The

[14]It is well known that if asked questions about similar issues all at once, respondents tend to make their answers consistent even when they would respond to the questions in substantially different ways were they asked separately.

coding is: 1=Not completed high school education, 2=High school graduate, 3=Some college, 4=Bachelor's degree or higher. Alternatively, we also use a set of binary indicator variables called HS DROPOUT, HIGH SCHOOL, SOME COLLEGE, BA DEGREE that are coded one if a respondent belongs to the respective category of EDUCATION and zero otherwise. Summary statistics for all variables used in the analysis are provided in appendix C.

6.4.2 Attitudes towards Highly and Low Skilled Immigrants and Natives' Skill Levels

Figure 6.3 plots the distributions of preferences conditional on respondents' skill levels. The results suggest two key findings. First, regardless of the respondents' skill level, highly skilled immigrants are strongly preferred over low skilled immigrants. Second, in stark contrast to the predictions based on the theoretical model, we find that support for both types of immigration is increasing (at a roughly similar rate) with respondents' skill level. For example, while only 7 percent of the least skilled respondents (those who did not finish high school) favor an increase in low skilled immigration, 29 percent favor an increase in highly skilled immigration. However, we find a similar preference differential among the most highly skilled respondents (those with at least a bachelor's degree): only 31 percent prefer an increase in low skilled immigration but more than 50 percent prefer an increase in highly skilled immigration.

Taken together these results are at odds with the claim that concerns about labor market competition are a driving force in shaping attitudes towards immigration. Instead, the results are consistent with previous findings indicating that people with levels of higher education are more likely to favor immigration (for a variety of other economic and non-economic reasons) regardless of immigrants' skill attributes (Hainmueller & Hiscox 2007).

6.4.3 Formal Tests of the Labor Market Competition Model

We created a binary indicator variable, HSKFRAME, coded one if the respondent i received the question about *highly skilled* immigrants and zero if he or she received the question about *low skilled* immigrants. The observed support for immigration is measured by the categorical variable PROIMIG which takes on the integer value associated with one of the five answer categories $j = (1, 2, ..., 5)$ from "strongly disagree" to "strongly agree" respectively. Let PROIMIG* be an unobserved continuous variable (with $-\infty \leq \text{PROIMIG}^* \leq \infty$) that represents a respondent's latent support for allowing in more immigrants. Both PROIMIG and PROIMIG* measure support for highly skilled or low skilled immigration depending on a respondent's assignment given by HSKFRAME. We model latent support for immigration by an ordered probit model:

$$\text{PROIMIG}_i^* \sim N(\mu_i, 1) \quad \text{where} \quad \mu_i = X_i\beta \tag{6.1}$$

and X_i is the row vector of observed respondent charateristics. The mapping of PROIMIG_i^* to PROIMIG_i is $\text{PROIMIG}_i = j$ if $\lambda_{j-1} \leq \text{PROIMIG}_i^* \leq \lambda_j$, with thresholds $\lambda_l \in \{(\lambda_0, ..., \lambda_5) | \lambda_0 = -\infty \wedge \lambda_5 = \infty \wedge \lambda_l < \lambda_{l+1}\} \; \forall \; l = (0, ..., 5)$. Thus, $p_{ij} = P(\text{PROIMIG}_i = j) = \Phi(\lambda_j|\mu_i) - \Phi(\lambda_{j-1}|\mu_i)$ where $\Phi(\cdot)$ is the normal cumulative density function with unit variance and mean μ_i. We estimate the coefficient vector β and the thresholds λ_1 through λ_{j-1} by maximising the log pseudolikelihood function:

$$ln\, L = \sum_{i=1}^{N} w_i \sum_{j=1}^{J} \mathbf{1}\{\text{PROIMIG}_i = j\} ln\, p_{ij} \tag{6.2}$$

where w_i is the poststratification weight and $\mathbf{1}\{\}$ is the indicator function. For all uncertainty estimates we employ the robust linearized variance estimator that yields the valid design based inferences.[15]

To explicitly test the labor market competition argument, we estimate the follow-

[15] Let $S(\beta) = \frac{\partial lnL}{\partial \beta}$ be the score function where $\hat{\beta}$ is estimated by solving $\hat{S}(\beta) = 0$. Following a first order Taylor series expansion the linearized variance estimator is then given by $\hat{V}(\hat{\beta}) = D\,V\{\hat{S}(\beta)\}|_{\beta=\hat{\beta}}D'$ where $D = \{\frac{\partial \hat{S}(\beta)}{\partial \beta}\}^{-1}$.

ing specification:

$$\begin{aligned}\mu_i &= \alpha + \gamma\,\text{HSKFRAME}_i + \delta\,(\text{HSKFRAME}_i \cdot \text{EDUCATION}_i) + \\ &\quad \theta\,\text{EDUCATION}_i + Z_i\psi\end{aligned} \tag{6.3}$$

where the parameter γ directly identifies the premium that natives attach to highly skilled immigrants relative to low skilled immigrants. If γ is positive this indicates that highly skilled immigrants are preferred relative to low skilled immigrants. If γ is negative this indicates that low skilled immigrants are preferred relative to highly skilled immigrants. The parameter δ captures how the premium for highly skilled immigration varies conditional on the skill level of the respondent.

The key predictions based on the standard model of labor market competition are as follows: For the least skilled respondents with $\text{EDUCATION}_i = 1$ (those who did not finish high school) we expect strong support for highly skilled over low skilled immigration so that $\gamma + \delta \cdot 1 > 0$. For the most highly skilled respondents with EDUCATION=4 (those with bachelor's degree or higher) we expect the exact opposite $\gamma + \delta \cdot 4 < 0$. In other words, low skilled immigration is preferred over highly skilled immigration. Taken together this implies that δ is negative, fairly large in magnitude ($|\gamma| > \delta/4$), and statistically significant.

In our second test specification we relax the assumption of linearity in the premium for highly skilled immigration and estimate:

$$\begin{aligned}\mu_i = \quad & \alpha + \gamma\,\text{HSKFRAME}_i + \sum_{k\in\{1,2,4\}} \delta_k\,(\text{HSKFRAME}_i \cdot \mathbf{1}\{\text{EDUCATION}_i = k\}) + \\ & \sum_{k\in\{1,2,4\}} \theta_k\,\mathbf{1}\{\text{EDUCATION}_i = k\} + Z_i\psi\end{aligned} \tag{6.4}$$

This specification allows for a different premium conditional on each of the four skill categories HS DROPOUT, HIGH SCHOOL, SOME COLLEGE, BA DEGREE. Notice that we use SOME COLLEGE (respondents that have some college education but did not graduate) as our reference category so that γ identifies the premium estimated for this skill level. Accordingly, $\gamma+\delta_1$, $\gamma+\delta_2$, and $\gamma+\delta_4$ identify the premia estimated for those respondents in the categories HS DROPOUT, HIGH SCHOOL, and BA DEGREE. The key prediction is that $\gamma + \delta_1$ is positive and significant while $\gamma + \delta_4$ should be negative and significant.

We also enter a basic set of socio-demographic covariates Z including the respondent's age (in 7 age brackets), gender (female=1,male=0), and race (four dummies for White, Hispanic, Black, and Other) in all specifications. The covariates are simply included here to increase the comparability of some of the coefficients with previous studies. Notice that since the randomization orthogonalizes HSKFRAME with respect to Z, the exact covariate choice does not affect the results of the main coefficients of interest.[16]

6.4.4 Results for Tests of the Labor Market Competition Model

Results for the tests are shown in Table 6.1. In the first two columns we separately regress attitudes towards highly skilled and low skilled immigration on respondents' skill level (measured by educational attainment) and the set of covariates. Following the labor market hypothesis we would expect that the support for low skilled (highly skilled) immigration should increase (decrease) in respondents' skill level. In contrast we find that the correlation between respondents' skill level and support for immigration is positive and significant for both types of immigration (columns one and two). In fact, we cannot reject the null hypothesis that the effect of respondents' skill on support for increased immigration is identical for highly skilled and low skilled immigrants (p=0.21).

The next three models implement our main experimental tests. To identify the premium attached to highly skilled relative to low skilled immigrants we use PROIMIG as our dependent variable and regress it on the indicator HSKFRAME that denotes whether a respondent received the frame about highly skilled immigrants rather than the question about low skilled immigrants. Results are shown in

[16]All results are substantively identical if additional (pre-treatment) covariates (like martial status, geographic indicators, etc.) or no covariates at all are used. Results available upon request.

column three. The high skill frame indicator enters positive and highly significant indicating that on average highly skilled immigrants are strongly preferred to low skilled immigrants. Column four includes the interaction of HSKFRAME with respondents' skill level, measured by EDUCATION. The interaction term enters with the expected negative sign, but it is statistically insignificant and the point estimate is very small in substantive terms. This result suggests that, in contrast to expectations based on the labor market competition model, the premium attached to highly skilled immigration does *not* vary significantly with respondents' skill level. In column five we also drop the linearity assumption regarding the effect of respondents' skill level and replace EDUCATION with our set of dummy variables that indicate the highest level of educational attainment (SOME COLLEGE is the reference category) plus all interactions with the high skill question frame. We find that not one of the interaction terms is significantly different from zero. A Wald test against the null that all interaction terms are jointly zero yields a p-value of (0.61) indicating that the variation in the premium attached to highly skilled immigration among differently skilled respondents is not significant.

Taken together these results reveal several striking features regarding the dynamic of respondents skill levels' and immigration preferences. In order to give some sense of the substantive magnitudes involved we simulate the predicted probability of supporting an increase in immigration (answers "somewhat agree" and "strongly agree" that the U.S. should allow more immigration) for the median respondent (white women aged 45) for all four skill levels and both immigration types based on the least restrictive model (model five in Table 6.1). Figure 6.4 shows the results and summarizes our key findings for the tests of the labor market competition argument.

First, in contrast to the predictions from the labor market competition model, support for both low and highly skilled immigration is steeply increasing in respondents' skill levels. This increase in the probability of supporting immigration is very large in substantive terms. For example, for highly skilled immigration it ranges from 0.23 [0.18; 0.26] among respondents who did not finish high school to 0.40 [0.35; 0.45] among college graduates (the numbers in square brackets give the .95 percent confidence interval). Furthermore, the increase is not linear, but instead is particularly pronounced for the gap between respondents who have a college education and those who do not. This plateau effect is in line with findings in some previous studies (Hainmueller & Hiscox 2007, Chandler & Tsai 2001) showing that exposure to university education seems to be the critical contributor to the generally positive relationship between education and support for immigration.

Second, regardless of the respondents' skill level, highly skilled immigrants are much preferred over low skilled immigrants. This finding is at odds with the expectation from the standard model of labor market competition that highly skilled natives should oppose inflows of highly skilled immigrants and support inflows of low skilled immigrants. On average (i.e., across the four skill levels), the predicted probability of supporting highly skilled immigration is about 0.15 higher than the probability of supporting low skilled immigration and this difference is highly statistically significant.

Third, there seems to be no systematic variation in the premium attached to highly skilled immigrants across respondents' skill level. As clearly indicated by the dashed lines that connect the predicted probabilities for each type of immigration, the step function that describes increased support for immigration with rising skill levels among respondents is quite similar for the two types of immigration. The relative differences in predicted probabilities of supporting highly skilled versus low skilled immigration are 0.17 [0.13; 0.20] for respondents who did not complete high school, 0.12 [0.10; 0.14] for high school graduates, 0.15 [0.12; 0.18] for those with some college education, and 0.17 [0.13; 0.21] for college graduates. The differences are not significantly different and do not have opposite signs, as predicted by the labour market competition model. The two dotted lines that connect the predicted support for the lowest and highest skill levels are almost exactly parallel in slope. This suggests that there is very little interaction between respondents' and immigrants' skill type in accounting for immigration preferences, and the results are sharply at odds from the expectation if labor market concerns were exercising a powerful influence – we would see a scissoring of these two slope lines. (The same figure based on our follow-up test that uses the within-group variation is presented in Figure 6.8 in appendix B. It mirrors the results obtained in Figure 6.4).

Finally, columns 7 and 8 in Table 6.1 present the results for the split sample tests that compare the relationship between respondents' skill levels and attitudes towards highly skilled and low skilled immigrants in the in-labor-force and the out-of-labour-force samples. Previous tests of the labor market competition model (Scheve & Slaughter 2001, Mayda 2006, Hainmueller & Hiscox 2007) have relied on similar tests, based on the idea that if labor market concerns are a driving factor in attitudes towards immigration, we should see marked differences across the two samples.[17] We find that the results are almost identical across the two sub-samples; in both cases highly skilled immigrants are preferred over their low skilled counterparts and this premium does not vary significantly with respondent skill level. This pattern is again inconsistent with what we would expect if concerns about labor market competition are shaping attitudes towards immigration.

6.5 Empirical Test II: The Fiscal Burden Model

6.5.1 Income and Fiscal Exposure to Immigration Across U.S. States

While the effect of immigration on natives via the labor market is modeled as a function of natives' skill levels, the impact of immigration via public finance is modeled as a function of natives' income (Facchini & Mayda 2007, Hanson et al. 2007). Progessitivity in tax systems means that richer natives pay more as a result of tax hikes (or benefit more from tax cuts) than do poorer natives; in addition, many types of public services and assistance are means-tested programs accessible only to poorer

[17] Consistent with these previous tests, our in-labour-force sample consists of full time, part time, and self employed respondents. The out-of-labour-force sample includes homemakers, retired, disabled, and other and those unemployed but looking for work. Alternative codings, such as including the unemployed in the in-labour force sample, leads to similar results (available upon request).

individuals. We construct a categorical variable called INCOME, which indicates a respondent's position in the income distribution. This coding ranges from 1 to 4 depending on whether a respondent belongs to the first, second, third, or fourth quartile of the household income distribution.[18]

The most prominent test of the fiscal burden model in the U.S. context focuses upon cross-state variation in exposure to the effects of immigration on government taxes and expenditures (Hanson et al. 2007). In the case of U.S. natives, exposure to the fiscal effects of immigration will depend upon the state in which they reside (and pay taxes). In general, the fiscal impact of immigrants should be strongest in states that have both a relatively large share of welfare reliant immigrants and a relatively generous welfare system. Hanson et al. (2007) note that measuring fiscal exposure to immigration is difficult, however, because individuals use public services in many forms: they use public safety, roads, parks, transportation, education, and healthcare, for example, as well as welfare programs. Furthermore, immigrants will contribute taxes and use public services and assistance to various degrees depending on both state policies and the characteristics of the immigrant population (e.g., income levels, family size, age, legal status, etc.). Hanson et al. (2007) construct two simple measures of fiscal exposure that focus only on state-level welfare spending, setting aside other types of spending and also immigrant tax contributions. In order to keep the analysis consistent with previous work we have reconstructed these two measures. We use the same data sources and the same coding approach, but employ data for the most recent years available (the 2006 American Community Service and the 2006 U.S. Census of Governments) to stay as close as possible to the time when our survey was fielded. Here we briefly describe the two measures.

Their first measure, FISCAL EXPOSURE I is equal to one for all states that meet two conditions: first, a high *total public welfare spending* per native household by state and local governments (states are coded as high welfare spending if they are

[18]Notice that our measure of household income has no missing data and is presumably fairly accurate because it is obtained from all KN panel members directly as part of the panel recruitment process.

above the national median on this measure);[19] second, a high ratio of immigrant to native households (states are code as high if they exceed the mean ratio across all the states). Our reconstructed measure for 2006 is summarized in the left panel of Figure 6.5 where the super-imposed dotted lines indicate the cutoff values for both axes. The nine states in the upper right corner are classified as having a high level of fiscal exposure to immigration.[20] Following Hanson et al. (2007) we also code a binary indicator called IMMIGHIGH that just identifies states that have a high ratio of immigrant to native households: IMMIGHIGH is coded as a one if the ratio is above the mean state immigrant-to-native household ratio and zero otherwise.[21] Notice that (by definition of FISCAL EXPOSURE I) all high fiscal exposure states are also high immigration states, but there are a few states with high immigration that are coded as having low fiscal exposure.

The second measure used by Hanson et al. (2007), FISCAL EXPOSURE II, is equal to one for states in which the *ratio of immigrant households receiving cash forms of public assistance* relative to the total number of native households exceeds a specified threshold (0.012).[22] Our measure for 2006 is reported in the right panel of

[19]We obtain *total public welfare spending* from the U.S. Census of Governments 2006. This measure accounts for most welfare benefits including cash, non-cash, and medical assistance. In particular, it includes expenditures associated with Supplemental Security Income (SSI), Temporary Assistance for Needy Families (TANF), Medicaid, food stamps, and expenditures for welfare activities not classified elsewhere. It excludes state spending on other public services such as public education, public safety, and public spaces. Unfortunately, welfare spending is not separately recorded for immigrant and native households, so one cannot isolate the amount of public assistance received by immigrants.

[20]Notice that we slightly refine the original measure by using the mean instead of the median welfare spending per native household as the cutoff for the first condition because it provides a much more natural cutoff in the data for 2006. All results are substantially identical if the median rate is used instead. Results available upon request

[21]In Figure 6.5 these are all states to the right of the dotted vertical line.

[22]We obtain the fraction of households receiving cash forms of public assistance from the 2006 American Community Survey. This measure is separately available for immigrant and native households, but limited to public assistance income including general assistance, TANF, and SSI. Compared to *total public welfare spending* it excludes non-cash benefits. Another disadvantage is that it merely measures the number of immigrant households receiving cash assistance but not the actual amount. We confirmed that both welfare spending measures are positively correlated (r=.28 in our data for 2006 compared to r=.24 in Hanson et al. (2007)). Based on this correlation, Hanson et al. (2007) argue that *total public welfare spending* serves as a reasonable proxy for welfare spending on

Figure 6.5. In our data, the exact same threshold (0.012) provides the natural break in the distribution. Seven states are marked as facing high fiscal pressure; Texas is the state just below the cutoff point. Overall the two fiscal exposure measures are highly correlated (r=.64) and agree on 46 of the 51 states. The classifications are almost identical to the ones originally used by Hanson et al. (2007).[23] Notice that conceptually FISCAL EXPOSURE I is the preferred measure of Hanson et al. (2007) because in contrast to FISCAL EXPOSURE II it includes non-cash benefits. However, it is important to note that FISCAL EXPOSURE I is based on a measure of general welfare spending and does not account for actual welfare uptake by immigrants and natives (captured, at least in part, by FISCAL EXPOSURE II).[24]

6.5.2 Attitudes towards Highly and Low Skilled Immigrants, Natives' Income, and Fiscal Exposure

Figure 6.6 plots the distribution of attitudes towards both highly skilled and low skilled immigration conditional on respondent income and the first measure of immigrant fiscal exposure (FISCAL EXPOSURE I).[25] To avoid a cluttering of the plot we focus on the fraction of respondents that is opposed to immigration (answers "somewhat disagree" and "strongly disagree" that the U.S. should allow more immigration) in each of the subsets defined by income, fiscal exposure, and immigration type. Two main findings emerge from the data.

immigrants only.

[23]For example, regarding FISCAL EXPOSURE II the list of states marked as high exposure differ by only a single state. For FISCAL EXPOSURE I all but six states seem to agree with certainty and for the others we cannot determine this with certainty as information on the cutoff values is missing in Hanson et al. (2007).

[24]Also notice that both measures are computed on a household basis, where an immigrant household is defined as one whose head was not a U.S. citizen at birth. This definition of immigrants includes foreign born naturalized citizens and U.S. born children of immigrants. The census and survey data do not distinguish the legal status of foreign-born respondents. This may affect public welfare measures, because in many states illegal immigrants are ineligible for most public services.

[25]The results are virtually identical if we use FISCAL EXPOSURE II instead.

First, the relationship between natives' income levels and attitudes towards highly skilled immigration is quite similar in high and low fiscal exposure states. Opposition to highly skilled immigration seems to be slightly lower among the richest natives than among the poorest natives, but this is true for both high and low fiscal exposure states. This finding seems at odds with what one would expect based upon the standard fiscal burden model, as opposition to highly skilled immigration should fall at a faster rate with income in those states that face a high fiscal exposure (given that highly skilled immigrants would relax the budget constraint through their net contributions to the tax coffers).

Second, the relationship between natives' income levels and attitudes towards low skilled immigration does vary dramatically between high and low fiscal exposure states. Opposition to low skilled immigration is increasing with respondent income in low exposure states, but the opposite is true for high fiscal exposure states where the richest natives are more welcoming of low skilled immigration than the poorest natives. This pattern is fundamentally inconsistent with the conventional wisdom that rich natives fear being burdened with higher taxes as a consequence of low skilled immigrants drawing on public services and draining government coffers. Instead, the results are more consistent with the alternative argument that poor natives should be the most vocal opponents of low skilled immigration in high fiscal exposure states because they fear increasing competition for public services and the erosion of welfare benefits.

6.5.3 Formal Tests of the Fiscal Burden Model

In order to formally test the fiscal burden model we estimate a series of ordered probit estimations with the following specification:

$$\mu_i = \alpha + \gamma\,\text{HSKFRAME}_i + \phi\,(\text{HSKFRAME}_i \cdot \text{INCOME}_i) + \tag{6.5}$$

$$\tau\,\text{INCOME}_i + Z_i\psi \tag{6.6}$$

where the parameter γ directly identifies the premium that natives attach to highly skilled over low skilled immigration and ϕ captures how the premium for highly skilled immigration varies conditional on the income level of the respondent. We estimate the models separately for the high and low fiscal exposure states and enter our basic set of socio-demographic covariates including the respondents age (in 7 age brackets), gender (female=1,male=0), and race (four dummies for White, Hispanic, Black, and Other) in all specifications. We also include respondent's education to each model, although the results are substantially identical if education is excluded.[26]

The key prediction from the standard fiscal burden model is that rich (high income) natives should attach a larger premium to highly skilled relative to low skilled immigrants than do poor natives. In addition, following Hanson et al. (2007), we can expect that this difference should be larger in states with high fiscal exposure to immigration than in states with low exposure. This means that we should expect ϕ to be positive and significant. In addition, ϕ should be larger in states with high fiscal exposure to immigrants than in low exposure states. This standard model assumes that taxes are adjusted to balance budgets, so natives attitudes reflect their concerns about tax rates. The alternative type of argument assumes that states adjust per capita transfers but hold tax rates constant, and makes an opposite prediction: rich natives prefer highly skilled over low skilled immigrants *less* than poor natives, and the difference should be larger in states with high fiscal exposure to immigrants than in states with low exposure. Accordingly we would instead expect ϕ to be negative and significant. And in the states with a high fiscal exposure to immigrants, ϕ should be larger (in absolute terms) than in low fiscal exposure states.

Finally, we also re-estimate all models relaxing the linearity assumption using the following specification:

$$\begin{aligned}\mu_i = \quad & \alpha + \gamma \text{HSKFRAME}_i + \sum_{k\in\{1,2,4\}} \phi_k \left(\text{HSKFRAME}_i \cdot \mathbf{1}\{\text{INCOME}_i = k\}\right) + \\ & \sum_{k\in\{1,2,4\}} \tau_k \mathbf{1}\{\text{INCOME}_i = k\} + Z_i\psi \end{aligned} \tag{6.7}$$

This specification allows for a different premium conditional on each of the four income quartiles, which we label INCOMEQ1 to INCOMEQ4. Notice that we use

[26]This is expected given the random assignment of HSKFRAME. Results available upon request.

INCOMEQ3 (respondents that fall in the third quartile of the income distribution) as our reference category.

6.5.4 Results for the Tests of the Fiscal Burden Model

The upper panel in Table 6.2 presents the estimation results. In the first column we estimate the model for all states. We find that income is associated with increased support for immigration. However, the premium attached to highly skilled over low skilled immigrants does not systematically vary across respondents' income levels; the coefficient for the interaction term between the high skill frame and the income variable enters insignificant and small in magnitude. In columns 2 and 3 we restrict the estimation to the sub-samples of states that are characterized by a high level of fiscal immigrant exposure according to the two measures FISCAL EXPOSURE I and II. Strikingly, the interaction terms in both subsamples now enter negative, fairly large in magnitude, and significant at conventional levels. This indicates that in high fiscal exposure states the premium attached to highly skilled immigration relative to low skilled immigration is decreasing in respondents' income level. This finding is clearly inconsistent with the conventional fiscal burden argument and the findings reported in previous studies. The finding is consistent, however, with the alternative argument according to which low income natives in high fiscal exposure states are likely to fear the erosion of welfare benefits as a consequence of low skilled immigration.

We find no such interaction in low fiscal exposure states, as shown in columns 4 and 5. In fact the interaction terms are almost exactly zero. In the last two columns 6 and 7 we further restrict the sub-samples to states with low fiscal exposure but high levels of immigration (as measured by the IMMIGHIGH variable). Again we find no interaction between the high skilled question frame and respondents' income levels. Taken together these two results suggest that the negative relationship between respondent income and the premium attached to highly skilled immigrants relative to low skilled immigrants is indeed driven by the levels of fiscal exposure and not by levels of immigration per se.

The lower panel of Table 6.2 presents the results when we also relax the linearity assumption in the interaction between respondents' income and preferences towards highly skilled and low skilled immigration. The findings are broadly similar to our previous findings, although they suggest that the main diving line in attitudes towards highly skilled relative to low skilled immigrants seems to be the transition from the second to the third quartile of the income distribution. In the high fiscal exposure states (models no 9 and 10) the interaction terms for both the lowest and the second lowest income dummies enter positive and with large magnitudes and are (jointly) highly significant, indicating that these two groups of respondents attach a larger premium to highly skilled relative to low skilled immigrants than respondents in the third quartile (the reference category). The interaction for the highest income dummy is almost zero indicating that for the richest respondents the premium is roughly similar to the premium for those in the third quartile. Again we find no such interaction in the states with low fiscal exposure (columns 11 and 12) and the subset of states with low exposure but high levels of immigration (columns 13 and 14).

In order to give a substantive interpretation to the results we simulate predicted probabilities for supporting an increase in immigration (answers "somewhat agree" and "strongly agree" that the US should allow more immigration) for the median respondent (white women aged 45 with some college education). We compute the predicted probabilities for all four income levels, both immigration types, and for both high and low fiscal exposure states (based on FISCAL EXPOSURE I) using our least restrictive models (models 10 and 12).[27] Figure 6.7 shows the results and summarizes our key findings regarding the fiscal burden model.

The figure suggests that the way in which fiscal concerns interact with respondents' income in forming attitudes towards highly skilled and low skilled immigrants is inconsistent with the conventional wisdom. In states with high fiscal exposure to immigration, poor respondents are less likely to support low skilled immigrants than they are in states with low exposure. Moreover, in high exposure states, poor natives attach a much larger premium to highly skilled relative to low skilled immigrants

[27]Results are substantivally identical if FISCAL EXPOSURE II is used instead. Results available upon request.

than they do in low exposure states. Taken together these results are much more consistent with the alternative type of argument about the fiscal concerns raised by immigration according to which poor natives fear competition with low skilled immigrants for access to public services. Rich respondents, meanwhile, are if anything more supportive of low skilled immigrants in high fiscal exposure states than they are states with low exposure, and the premium they attach to highly skilled immigrants is unaffected by fiscal exposure, findings that are completely at odds with the standard argument that rich natives are primarily concerned about tax hikes that could be triggered by low skilled immigration. (The same figure based on our follow-up test that uses the within-group variation is presented in Figure 6.9 in appendix B. It mirrors the results obtained in Figure 6.7).

6.6 Conclusion

To date, no empirical study has been able to distinguish between the attitudes that native citizens have towards highly skilled immigrants and their attitudes towards low skilled immigrants. This distinction is a critical feature of the theoretical models that link economic concerns with attitude formation and policy preferences with respect to immigration. In our survey experiment we were able to explicitly and separately examine individuals' attitudes towards highly skilled and low skilled immigrants, randomly assigning respondents to answer questions about immigrants with different skill levels.

The results from the survey experiment challenge the predictions made by the standard theoretical models and the conclusions reached in recent (non-experimental) studies. The labor market competition model predicts that natives will be most opposed to immigrants who have similar skill levels to their own. We find instead that *both* highly skilled and low skilled respondents strongly prefer highly skilled immigrants over low skilled immigrants, and this preference is not decreasing in respondents' skill levels. Support for *both* highly skilled and low skilled immigration is strongly increasing in respondents' skill levels. We also find that these relationships are similar for respondents currently in or currently out of the labor force. Overall, the results indicate that, in general, concerns about labor market competition are not a powerful driver of anti-immigrant sentiment in the United States – or, at least, not in the simple ways so far imagined.

According to the standard fiscal burden model, rich natives oppose low skilled immigration more than do poor natives, and this difference should be larger in states with greater fiscal exposure in terms of immigrant access to public services. We find instead that rich and poor natives are *equally* opposed to low skilled immigration in general, and rich natives are actually less opposed to low skilled immigration in high exposure states than in low exposure states. These results are clearly inconsistent with claims that concerns about a heavier tax burden associated with the provision of public services are driving rich natives to oppose low skilled immigration.

We find evidence, however, that supports an alternative argument about public finance and immigration. We find that poor natives are markedly more opposed to low skilled immigration in states with high fiscal exposure than in states with low fiscal exposure. This supports an argument that concerns about access to or overcrowding of public services contribute to anti-immigrant attitudes among poorer citizens. Across the states over the past 25 years or so, while immigration has had no discernable impact on tax rates, per capita welfare expenditures have grown the slowest in states that experienced larger increases in the share of immigrants in their population. The evidence suggests that fears among poor natives about constraints on welfare benefits as a result of immigration are far more relevant than concerns among the rich about increased taxes.

6.7 Appendix A: Theoretical Framework

We incorporate a simple model of public finance with the standard factor-proportions (FP) analysis of immigration to derive the basic propositions about natives' attitudes towards highly skilled and low skilled immigrants. We build on similar analysis by Dustman and Preston (2006) and Facchini and Mayda (2007), and where possible use matching notation. Assume a nondiversified economy producing one commodity, with constant returns to scale, using two factors of production: highly skilled labor (L_S) and low skilled labor (L_U). The native population is made up of $N = L_S + L_U$ individuals, each owning one unit of labor (either highly skilled or low skilled) and an endowment e^n of the commodity (where n indexes natives). Equilibrium is described by full employment of each factor and competitive profits:

$$a_S Q = L_S \tag{6.8}$$

$$a_U Q = L_U \tag{6.9}$$

$$a_S w_S + a_U w_U = 1 \tag{6.10}$$

where a_S and a_U are the quantities of each factor required per unit of output Q, w_S and w_U are wages for highly skilled and low skilled labor, and the commodity price is fixed in the world market and normalized to 1. After total differentiation, given cost minimizing values for a_S and a_U, we can derive solutions that express changes in wages as a function of different types of immigration:

$$\hat{w}_S = \frac{(1-\theta_S)}{\sigma}\left(\hat{L}_U - \hat{L}_S\right) \tag{6.11}$$

$$\hat{w}_U = -\frac{(1-\theta_U)}{\sigma}\left(\hat{L}_U - \hat{L}_S\right) \tag{6.12}$$

where hats indicate proportional changes, θ_j is the distributive share of L_j in total output ($j \in \{S, U\}$), and σ is the elasticity of substitution between factors. It is clear that any increase in the supply of highly skilled labor ($\hat{L}_S > 0$), ceteris paribus, implies a reduction in real wages for highly skilled natives ($\hat{w}_S < 0$) and a rise in real wages for low skilled natives ($\hat{w}_U > 0$). Alternatively, inflows of low skilled labor ($\hat{L}_U > 0$), ceteris paribus, will raise real wages of highly skilled natives ($\hat{w}_S > 0$) and reduce real wages of low skilled natives ($\hat{w}_U < 0$). These are the two scenarios presented in the survey experiment. Of course, if there are inflows of both highly skilled and low skilled immigrants, the wage effects will depend on the impact of the inflows on relative factor supplies ($\hat{L}_S - \hat{L}_U$).

Assume that the government provides public services to all individuals residing in the country and that these services are consumed in equal amounts by all and valued at b per person (so that they are, in effect, a lump sum transfer of b to each resident). Government spending is financed by a proportional income tax, set at rate τ, so that the government budget constraint is:

$$\tau\left(w_S L_S + w_U L_U + E\right) = b\left(L_S + L_U\right) \tag{6.13}$$

where $E = \sum e^n$. The after-tax income of the n-th native is:

$$I_j^n = (1-\tau)\left(w_j + e^n\right) + b \tag{6.14}$$

Immigration can affect the after-tax income of a native by altering wage rates, but also by affecting the tax rate or the provision of government services (or both).

In line with previous approaches, we assume that the government will adjust to any change in fiscal circumstances by *either* adjusting the tax rate *or* by adjusting spending. In the first case, holding b constant and totally differentiating equation 6.13 yields:

$$\hat{\tau} = (\lambda_S - \phi_S)\,\hat{L}_S + (\lambda_U - \phi_U)\,\hat{L}_U - \phi_E \hat{E} \tag{6.15}$$

where λ_j is the share of L_j in the population and ϕ_j is the distributive share of L_j in total income ($Q + E$). Assuming $w_S > w_U$, then $\lambda_U - \phi_U > 0$ and it is clear that inflows of low skilled immigrants ($\hat{L}_U > 0$) necessitate raising the tax rate, all else equal, as taxes on their wages (at the current rate) will not cover the additional spending on the government services they consume. It is possible that such immigrants could arrive with endowments ($\hat{E} > 0$) enough to generate an offsetting increase in tax revenues, but the standard assumption is that low skilled immigrants have zero taxable assets. The arrival of highly skilled immigrants ($\hat{L}_S > 0$) will lead to a reduction in the tax rate, all else equal, if $\lambda_S - \phi_S < 0$, which is the case when $E < L_U(w_S - w_U)$. The intuition here is that highly skilled immigrants will raise per capita before-tax income, which at the fixed levels of per capita government spending allows a reduction in the tax rate (as long as endowments do not represent a large proportion of national income). This tax relief affect is accentuated to the extent that highly skilled immigrants bring taxable endowments.

After totally differentiating equation 6.14, we can describe the impact of immigration on native n's after-tax income:

$$\hat{I}_j^n = \frac{w_j(1-\tau)\hat{w}_j - \tau G_j^n \hat{\tau}}{(1-\tau)G_j^n + b} \tag{6.16}$$

where gross (before-tax) income $G_j^n = (w_j + e^n)$. What can we now say about the impact of different types of immigration on the net income of natives? Holding aside the wage effect, which we know (from equations 6.11 and 6.12 above) will hinge on the skill level of the particular native, it is easy to see that the impact will vary with income. Combining 6.15 and 6.16, and assuming for simplicity that $\hat{E} = 0$, it is straightforward to show that with inflows of low skilled immigrants ($\hat{L}_U > 0$), the tax rate must rise ($\hat{\tau} > 0$), net incomes fall ($\hat{I}_j^n < 0$), and the losses are magnified for natives with higher gross incomes ($\partial \hat{I}_j^n / \partial G_j^n < 0$). Conversely, with inflows of highly skilled immigrants ($\hat{L}_S > 0$), the tax rate falls ($\hat{\tau} < 0$) as long as $E < L_U(w_S - w_U)$, net incomes rise ($\hat{I}_j^n > 0$), and the gains are greater for those with higher before-tax incomes ($\partial \hat{I}_j^n / \partial G_j^n > 0$). In sum, richer natives lose more than poorer counterparts from the entry of low skilled immigrants, and they gain more with the arrival of highly skilled immigrants.

The overall effect of immigration on the net income of native n, with skill level j, will depend on the combination of wage and tax effects. For low skilled natives, these effects are always in the same direction: inflows of low skilled immigrants will reduce wages ($\hat{w}_U < 0$) and raises taxes, while inflows of highly skilled workers raises wages ($\hat{w}_U > 0$) and reduces taxes. Highly skilled natives have a more complicated calculation: low skilled immigrants raise their wages ($\hat{w}_S > 0$) but also increase the tax burden; highly skilled immigrants push down wages ($\hat{w}_S < 0$) but also decrease taxes.

What if the government adjusts to the change in fiscal circumstances by adjusting spending while keeping the tax rate fixed? In this second case, holding τ constant and totally differentiating equation 6.13 yields:

$$\hat{b} = -\left(\lambda_S - \phi_S\right)\hat{L}_S - \left(\lambda_U - \phi_U\right)\hat{L}_U + \phi_E \hat{E} \tag{6.17}$$

The impact of immigration on the per-capita provision of government services when taxes are fixed is just the exact reverse of the effect on the tax rate when spending is fixed. Inflows of low skilled immigrants ($\hat{L}_U > 0$) necessitate a reduction in per-person services ($\hat{b} < 0$), assuming such immigrants bring no taxable endowments. Highly skilled immigrants ($\hat{L}_S > 0$) generate an expansion in services ($\hat{b} > 0$).

Totally differentiating equation 6.14, this time assuming no change in the tax rate but an adjustment in spending, we get:

$$\hat{I}_j^n = \frac{w_j(1-\tau)\hat{w}_j + b\hat{b}}{(1-\tau)G_j^n + b} \tag{6.18}$$

Controlling for the wage effect, and assuming $\hat{E} = 0$, it is easy to show that with inflows of low skilled immigrants ($\hat{L}_U > 0$) per-capita services must be cut ($\hat{b} < 0$) and net incomes fall ($\hat{I}_j^n < 0$); these losses are smaller for natives with higher gross incomes ($\partial\hat{I}_j^n/\partial G_j^n > 0$). Inflows of highly skilled immigrants ($\hat{L}_S > 0$) result in an expansion of services ($\hat{b} > 0$) and an increase in net incomes ($\hat{I}_j^n > 0$), but these gains are smaller for those with higher incomes ($\partial\hat{I}_j^n/\partial G_j^n < 0$). In this case, the stakes are largest for the poorest natives: poor natives are hurt more than richer natives by low skilled immigration, and they benefit more than richer counterparts from highly skilled immigration.

6.8 Appendix B: Within-Groups Analysis

This appendix summarizes the results from the within-groups analysis of immigration attitudes. Two weeks following the implementation of the main survey (module

1) we contacted a randomly chosen subset of 1,400 of the respondents with a second survey (module 2). In module 2 we asked respondents about their attitudes towards immigration using the same two question versions that we used in module 1, one referring to *highly skilled* immigration and the other referring to *low skilled* immigration. We randomly allocated the module 2 questions among the respondents according to a split-sample cross-over design: Of the respondents that received the *highly skilled* immigration question in module 1, half received the *highly skilled* immigration question and half received the *low skilled* immigration question in module 2. Similarly, of the respondents that received the *low skilled* immigration question in module 1, half received the *highly skilled* immigration question and half received the *low skilled* immigration question in module 2. The allocation and frequencies are summarized in Table 6.3.

This design allows us to examine both the stability of attitudes over time as well as the degree to which the same respondents prefer highly skilled versus low skilled immigration. Since the groups are randomly assigned in both modules, each group comparison – within and across modules – in principle provides unbiased estimates. Notice that in contrast to the between-group comparisons for module 1 that are the focus in the main text, the across-module within-group comparisons that we focus on in this appendix may be affected by carry-over effects if the respondents' answer to the module 1 question affects her answer to the module 2 question. However, given the two week "wash out" period between the modules this seems unlikely (see tests in the next section).

6.8.1 Stability of Attitudes and Skill Premium

Table 6.4 summarizes the mean support for the two types of immigration for the subset of respondents that participated in both module 1 and module 2. The observed support for immigration in each module is measured by the categorical variable PROIMIG which takes on the integer value associated with one of the five answer categories $j = (1, 2, ..., 5)$ from "strongly disagree" to "strongly agree" respectively. We find that attitudes towards both highly skilled and low skilled immigration are fairly stable between module 1 and module 2. The first row suggests that among those who are asked about low skilled immigration in both modules, we cannot reject the null that the mean level of support is the same in both modules (using paired-tests with sampling weights). The second row suggests that the same is true for those who are asked about highly skilled immigration in both modules.

The third row looks at the within-groups differences for those who were asked about low skilled immigration in module 1 and highly skilled immigration in module 2. We find that high skilled immigrants are much preferred over their low skilled counterparts: on average the support is about .5 higher (on the 5 point scale) and this difference is highly significant. Row four indicates that the same is true when the question order is reversed (highly skilled in module 1 and low skilled in module 2): Support is about .66 higher and the difference is highly significant. We cannot reject the null that the premium attached to highly skilled versus low skilled immigration differs depending on the question order (the confidence intervals of the two differences in row 3 and 4 overlap widely). This suggests that the two week "wash out" period between the two survey modules was sufficient to eliminate carry-over effects that may result from the fact that respondents try to make their answers consistent to avoid the impression of skill-based discrimination between the two types of immigrants.

6.8.2 Tests of the Labor Market Competition Model

To test the labor market competition model based on within-groups differences we focus on the 673 respondents that were asked about different skill types in module 1 and 2. We fit ordered probit models as described in the main text, regressing attitudes toward highly skilled and low skilled immigration on a set of highest educational attainment dummies and our basic set of covariates. Again, we simulate the predicted probability of supporting an increase in immigration (answers "somewhat agree" and "strongly agree" that the U.S. should allow more immigration) for the median respondent (white women aged 45) for all four skill levels and both immigration types. Figure 6.8 shows the results and summarizes our key findings for the tests of the labor market competition argument based on the within-groups analysis. Full regression results are available upon request.

The findings are virtually identical to the results obtained from the between-groups test presented in the main text, except that the confidence intervals are slightly larger due to the fact that our sample size is now cut in half. In contrast to the predictions from the labor market competition model, support for both low and highly skilled immigration is steeply increasing in respondents' skill levels. Moreover, there seems to be no systematic variation in the premium attached to highly skilled immigrants across respondents' skill level.

We also replicated the tests comparing the support for highly skilled and low skilled immigration among those respondents that are in and out of the labour force. The results are very similar to the between-groups test reported in the paper. We find that when going from lowest to the highest educational attainment category the probability of supporting an increase in highly skilled immigration increases by .25 [.17; .33] among those that are currently in the labour force and by .30 [.22 ; .39] among those that are out of the labour force. Similarly, the probability of supporting an increase in low skilled immigration raises by .18 [.13 ; .22] among those in the labour force and by 0.12 [0.08; 0.17] among those out of the labour force. These results are again inconsistent with the idea that labor market concerns are a driving factor in attitudes towards immigration. If that where the case, we should see marked differences across the two labor market sub-samples. Full regression results are available upon request.

6.8.3 Tests of The Fiscal Burden Model

To test the fiscal burden model based on within-groups differences we again focus on the 673 respondents that were asked about different skill types in module 1 and 2. For both high and low fiscal exposure states we fit ordered probit models as described in the main text, regressing attitudes toward highly skilled and low skilled immigration on a set of dummies that indicate a respondent's quartile in the income distribution, our basic set of covariates, and educational attainment. Again, we simulate the predicted probability of supporting an increase in immigration (answers "somewhat agree" and "strongly agree" that the U.S. should allow more immigration) for the median respondent (white women aged 45) for all four skill levels, both immigration types, and fiscal exposure levels. Figure 6.9 shows the results (based on FISCAL EXPOSURE I) and summarizes our key findings for the tests of the fiscal burden model based on the within-groups analysis. The results are substantivally identical if FISCAL EXPOSURE II is used instead. Full regression results are available upon request.

The findings are again very similar to the results reported in the main text based on the between-groups analysis. In high fiscal exposure states, the premium attached to high versus low skilled immigrations is small among the richest natives, but large among the poorest natives. This indicates that tax concerns are unlikely to be an important driver of anti-immigrant sentiments because in states with high exposure the richest natives should be the ones that attach the highest premium to high skilled immigration. The results are supportive of the alternative scenario that highlights fears about the erosion of public services. In high exposure states where competition for public services is most serve, the poorest natives exhibit the highest premium for highly skilled versus low skilled immigrants. Similarly, comparing low and high exposure states the results are inconsistent with the tax hike argument. Rich (poor) respondents, meanwhile, are if anything more supportive of low (high) skilled immigrants in high fiscal exposure states than they are states with low exposure. These findings are again completely at odds with the standard argument that rich natives are primarily concerned about tax hikes that could be triggered by low skilled immigration. The findings support the alterative scenario which anticipates that fear of competition with low skilled immigrants for access to public services, especially among the poorest natives is critical for anti-immigrant sentiment.

6.9 Appendix C: Descriptive Statistics

variable	obs	mean	sd	min	max
PROIMIG	1589	2.57	1.25	1	5
FEMALE	1601	0.51	0.50	0	1
WHITE	1601	0.73	0.45	0	1
BLACK	1601	0.10	0.30	0	1
HISPANIC	1601	0.03	0.17	0	1
AGE CATEGORY	1601	3.85	1.68	1	7
HSKFRAME	1601	0.50	0.50	0	1
EDUCATION	1601	2.76	1.00	1	4
HS DROPOUT	1601	0.11	0.31	0	1
HIGH SCHOOL	1601	0.32	0.47	0	1
SOME COLLEGE	1601	0.28	0.45	0	1
BA DEGREE	1601	0.30	0.46	0	1
INCOME	1601	2.54	1.11	1	4
INCOMEQ1	1601	0.25	0.43	0	1
INCOMEQ2	1601	0.20	0.40	0	1
INCOMEQ3	1601	0.31	0.46	0	1
INCOMEQ4	1601	0.24	0.43	0	1
FISCAL EXPOSURE I	1601	0.25	0.43	0	1
FISCAL EXPOSURE II	1601	0.30	0.46	0	1
IMMIGHIGH	1601	0.52	0.50	0	1

6.10 Tables for Chapter 6

Table 6.1: Individual Support for Highly Skilled and Low Skilled Immigration - Test of the Labour Market Competition Model

Dependent Variable	In Favor of: High Skilled Immig.	Low Skilled Immig.	In Favor of: Immigration				
Model No	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sub-sample						labour force in	out
EDUCATION	0.21	0.27		0.27		0.33	0.19
	(0.05)	(0.05)		(0.05)		(0.06)	(0.07)
HSKFRAME			0.54	0.73	0.56	0.73	0.64
			(0.07)	(0.20)	(0.12)	(0.28)	(.29)
HSKFRAME·EDUCATION				-0.07		-0.08	0.00
				(0.07)		(0.09)	(0.11)
HS DROPOUT					-0.41		
					(0.18)		
HSKFRAME·HS DROPOUT					0.24		
					(0.25)		
HIGH SCHOOL					-0.16		
					(0.12)		
HSKFRAME·HIGH SCHOOL					-0.05		
					(0.17)		
BA DEGREE					0.41		
					(0.12)		
HSKFRAME·BA DEGREE					-0.08		
					(0.16)		
N	798	791	1589	1589	1589	946	643
Covariates	x	x	x	x	x	x	x

Note: Order Probit Coefficients shown with standard errors in parenthesis. All models include a set of the covariates age, gender, and race (coefficients not shown here). The reference category for the set of education dummies is SOME COLLEGE (respondents with some college education).

Table 6.2: Individual Support Highly Skilled and Low Skilled Immigration - Test of the Fiscal Burden Model

Dependent Variable	In Favor of Increasing Immigration						
Level of Fiscal Exposure	Both	High		Low		Low	
Measure of Fiscal Exposure		I	II	I	II	I	II
Level of Immigration	Both	Both		Both		High	
Model No	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HSKFRAME	0.62	1.22	1.18	0.43	0.44	0.33	0.39
	(0.17)	(0.34)	(0.29)	(0.19)	(0.20)	(0.31)	(0.34)
HSKFRAME·INCOME	-0.03	-0.18	-0.20	0.03	0.04	0.05	0.08
	(0.06)	(0.12)	(0.10)	(0.07)	(0.07)	(0.11)	(0.12)
INCOME	0.11	0.22	0.26	-0.03	-0.05	0.02	-0.01
	(0.04)	(0.08)	(0.08)	(0.05)	(0.05)	(0.09)	(0.09)
N	1589	397	470	1192	1119	431	358
Covariates	x	x	x	x	x	x	x
Model No	(8)	(9)	(10)	(11)	(12)	(13)	(14)
HSKFRAME	0.57	0.54	0.46	0.60	0.64	0.54	0.72
	(0.11)	(0.23)	(0.22)	(0.13)	(0.13)	(0.25)	(0.27)
INCOMEQ1	0.17	-0.39	-0.45	0.35	0.36	0.11	0.20
	(0.13)	(0.26)	(0.26)	(0.14)	(0.15)	(0.27)	(0.27)
HSKFRAME·INCOMEQ1	-0.14	0.39	0.39	-0.30	-0.30	-0.00	0.05
	(0.18)	(0.37)	(0.33)	(0.20)	(0.21)	(0.35)	(0.39)
INCOMEQ2	-0.03	-0.52	-0.44	0.12	0.16	0.22	0.39
	(0.14)	(0.29)	(0.24)	(0.16)	(0.16)	(0.22)	(0.26)
HSKFRAME·INCOMEQ2	0.30	0.69	0.65	0.17	0.12	-0.36	-0.62
	(0.18)	(0.36)	(0.32)	(0.21)	(0.22)	(0.34)	(0.38)
INCOMEQ4	0.34	0.24	0.28	0.33	0.33	0.24	0.26
	(0.12)	(0.23)	(0.21)	(0.14)	(0.15)	(0.23)	(0.27)
HSKFRAME·INCOMEQ4	-0.18	-0.04	-0.08	-0.20	-0.20	0.06	0.04
	(0.17)	(0.31)	(0.30)	(0.20)	(0.20)	(0.33)	(0.36)
N	1589	397	470	1192	1119	431	358
Covariates	x	x	x	x	x	x	x

Note: **Order Probit Coefficients shown with standard errors in parenthesis. All models include a set of the covariates age, gender, and race, and education (coefficients not shown here). The reference category for the set of income quartile dummies is INCOMEQ3 (respondents in the third quartile of the income distribution).**

Table 6.3: Split-Sample Cross-Over Design for Within-Groups Test

Module 1		**Module 2**	
Question frame:	N	**Question frame:**	N
highly skilled immigrants	798	*highly skilled* immigrants	342
		low skilled immigrants	337
low skilled immigrants	791	*highly skilled* immigrants	338
		low skilled immigrants	339
Total	1,589	Total	1,356

Table 6.4: Mean Support for Immigration by Module and Immigrants' Skill Type

Comparison			**PROIMIG**				
Module 1	**Module 2**	**N**	**Mod 1**	**Mod 2**	**Dif**	**.95 LB**	**.95 UB**
Low skilled	Low skilled	330	2.29	2.21	0.08	-0.04	0.19
Highly skilled	Highly skilled	341	2.79	2.88	-0.09	-0.21	0.03
Low skilled	Highly skilled	337	2.17	2.66	-0.49	-0.65	-0.34
Highly skilled	Low skilled	336	2.86	2.20	0.66	0.52	0.81

6.11 Figures for Chapter 6

Figure 6.1: Changes in Average Marginal State Income Tax Rate, Public Welfare Spending Per Capita, and Percent Foreign Born Population: 2004 to 1990

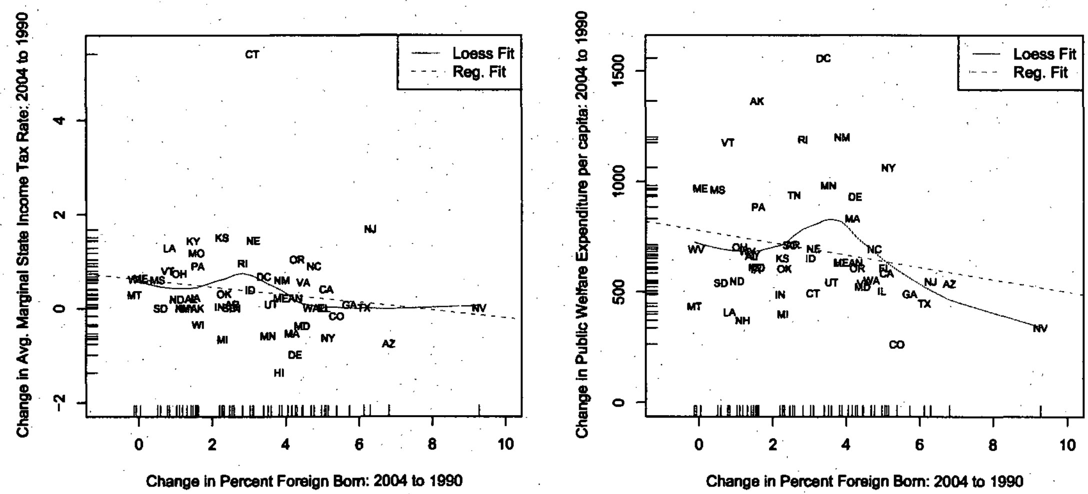


Figure 6.2: Support for Highly Skilled and Low Skilled Immigration

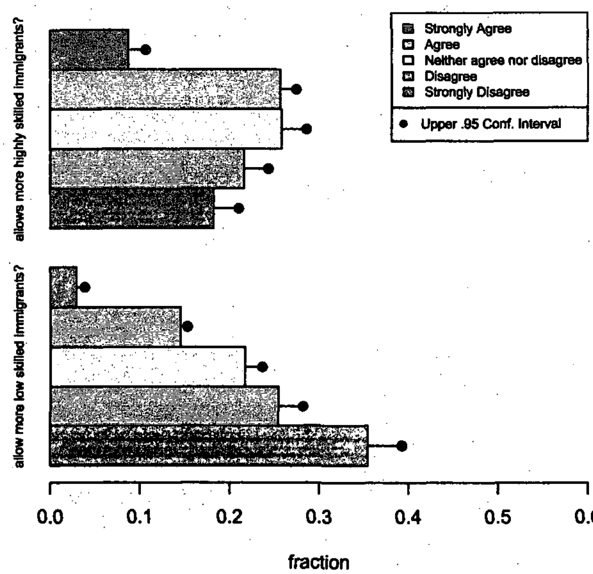


Figure 6.3: Support for Highly Skilled and Low Skilled Immigration by Respondents' Skill Level

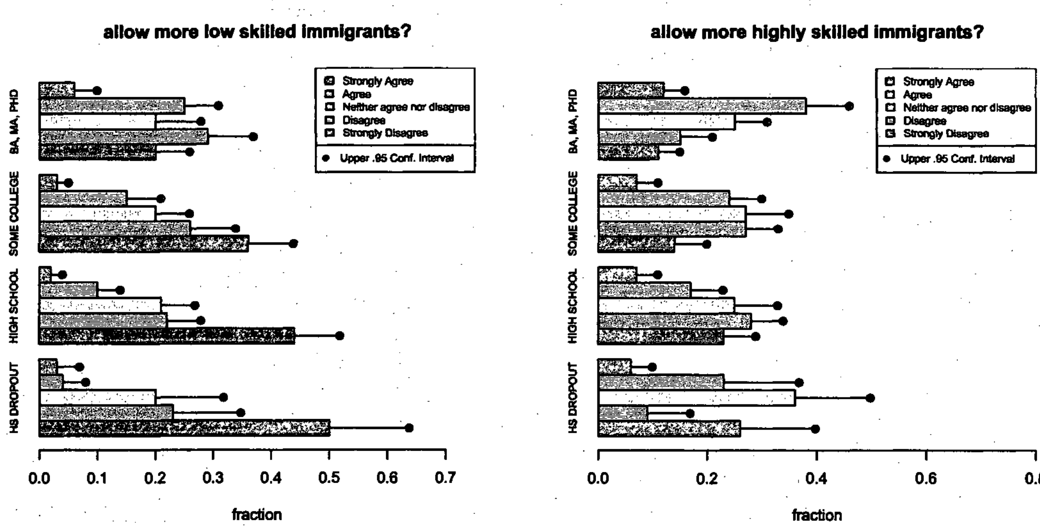


Figure 6.4: Support for Highly Skilled and Low Skilled Immigration by Respondents' Skill Level

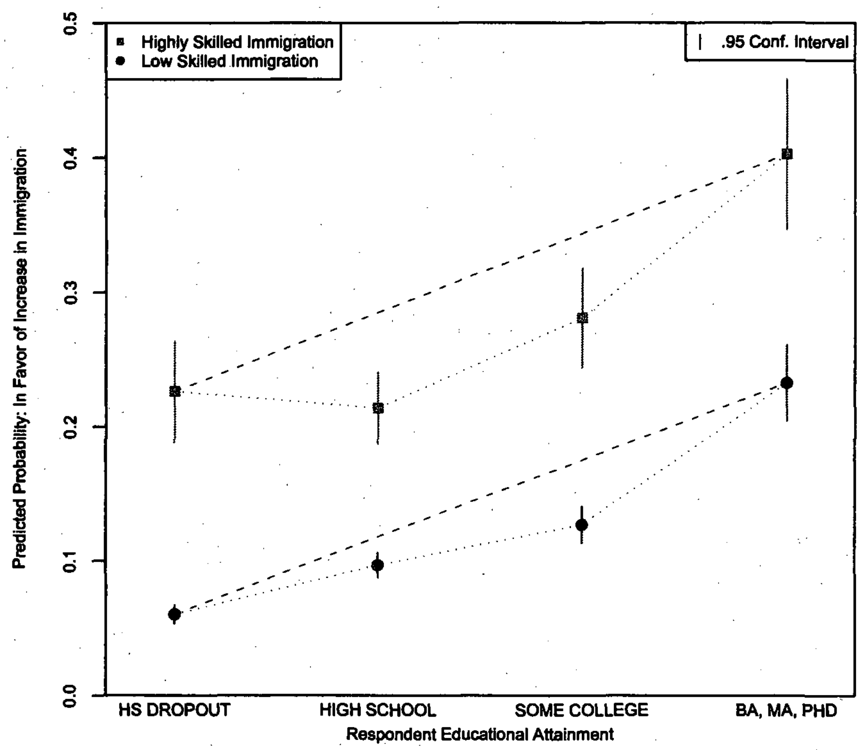


Figure 6.5: Measures of Fiscal Exposure

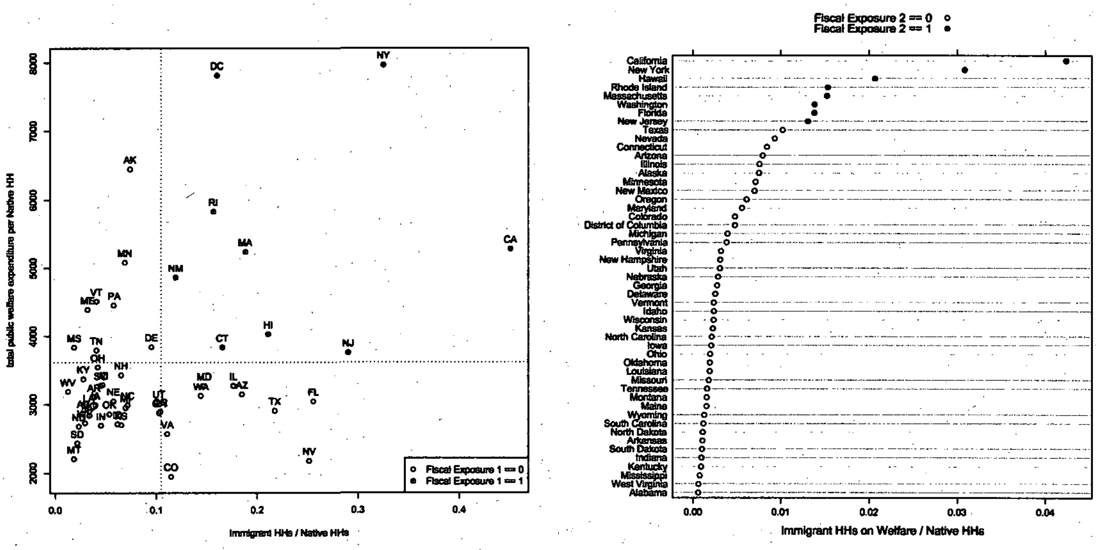
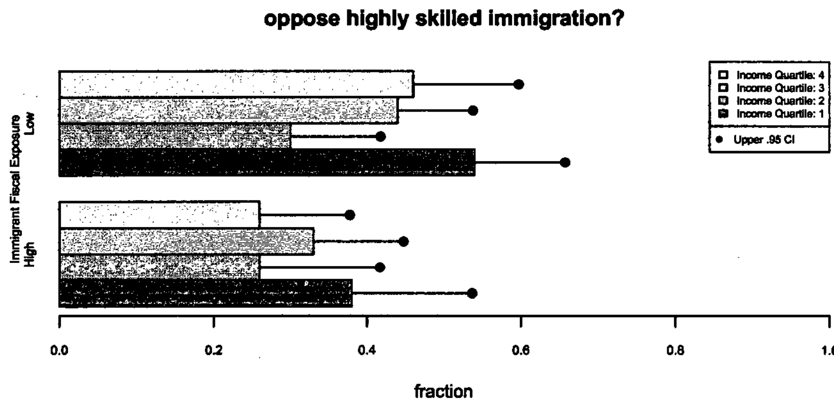


Figure 6.6: Attitudes Toward Highly Skilled and Low Skilled Immigration by Respondents' Income Level and Immigrant Fiscal Exposure of Respondents' State



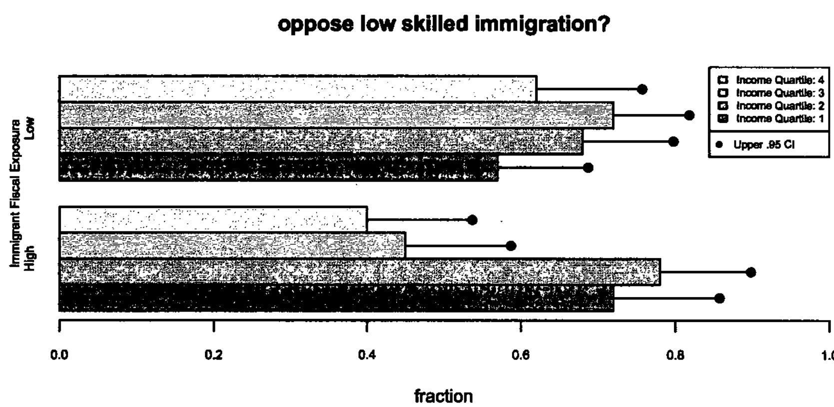


Figure 6.7: Support for Highly Skilled and Low Skilled Immigration by Respondents' Income Level and Immigrant Fiscal Exposure of Respondents' State

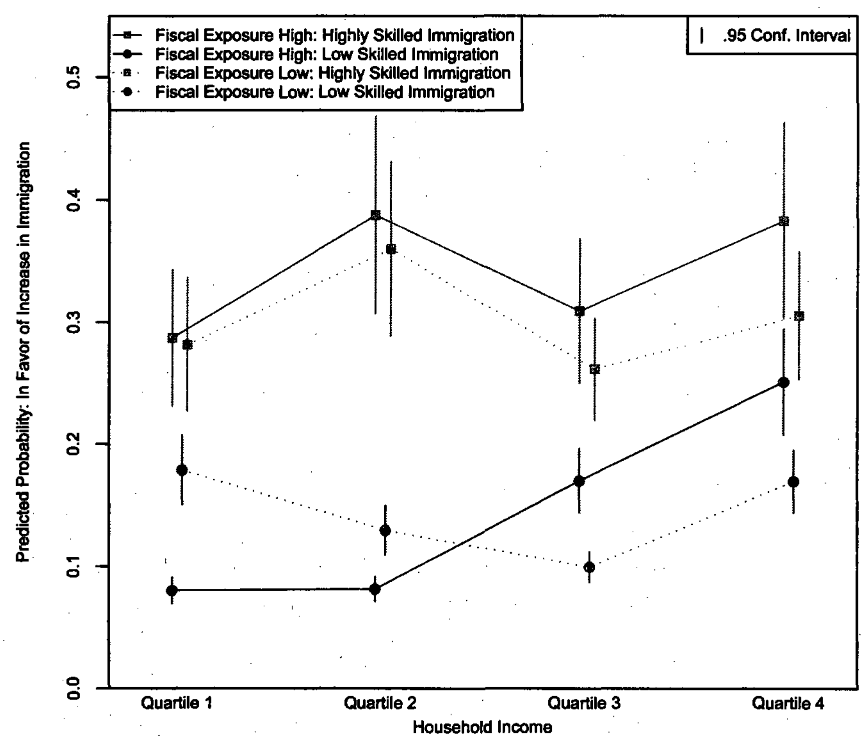


Figure 6.8: Within-Groups Test: Support for Immigration by Respondents' Skill Level

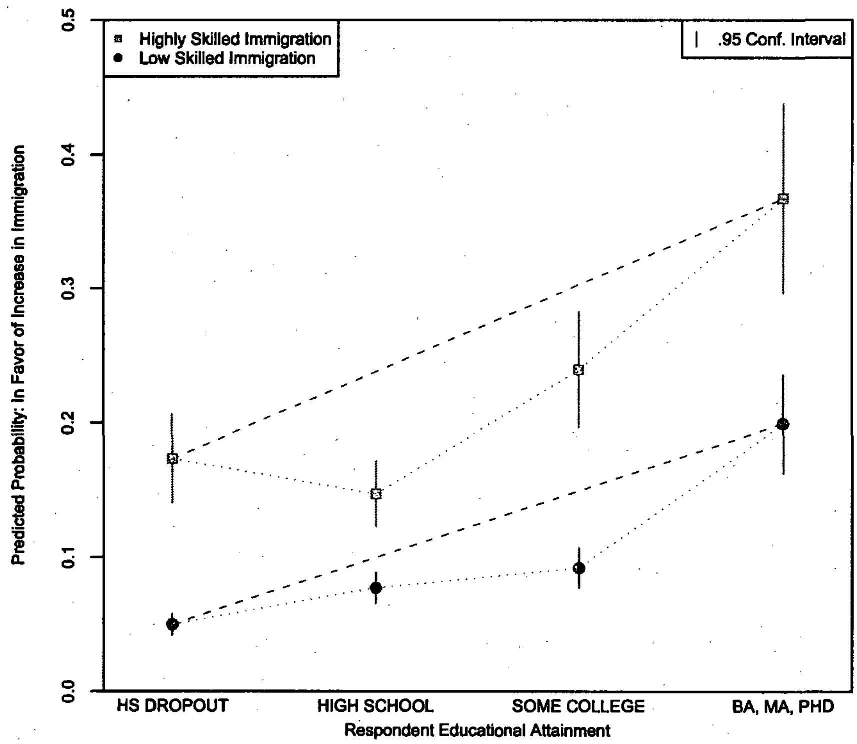
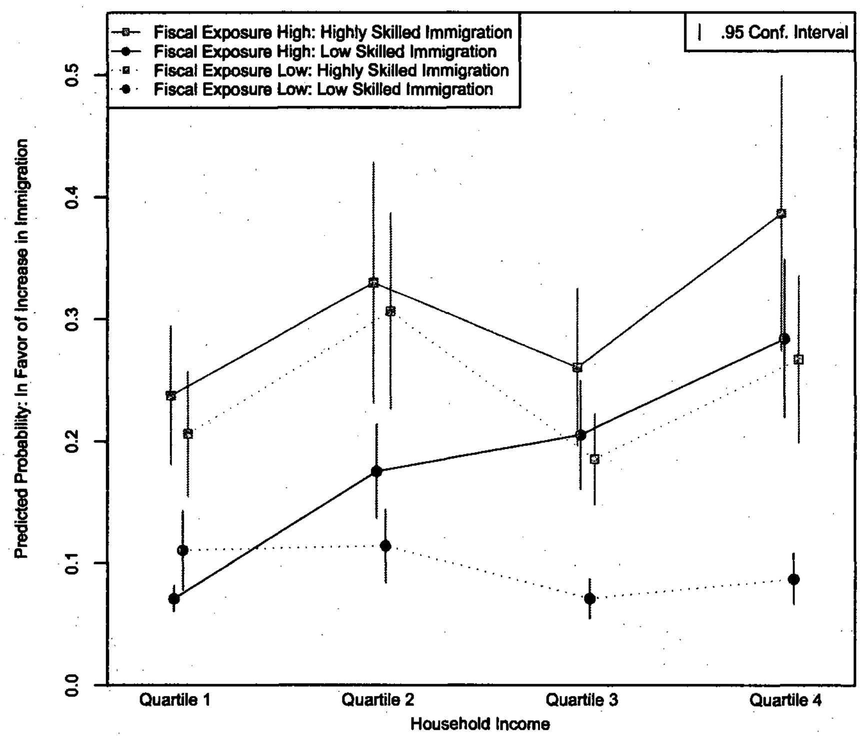


Figure 6.9: Within-Groups Test: Support for Immigration by Respondents' Income Level



Bibliography

(n.d.).

Abadie, A. (2003), 'Semiparametric instrumental variable estimation of treatment response models', *Journal of Econometrics* **113**(2), 231–263.

Abadie, A., Diamond, A. & Hainmueller, J. (2009), 'Synth: An r package for synthetic control methods in comparative case studies', *Journal of Statistical Software* **Forthcoming**.

Abadie, A. & Gardeazabal, J. (2003), 'The Economic Costs of Conflict: A Case Study of the Basque Country', *The American Economic Review* **93**(1), 113–132.

Abadie, A. & Imbens, G. (2006), 'Large Sample Properties of Matching Estimators for Average Treatment Effects', *Econometrica* **74**(2), 235–267.

Abadie, A. & Imbens, G. (2007), 'Simple and Bias-Corrected Matching Estimators for Average Treatment Effects'.

Angrist, J., Imbens, G. & Rubin, D. (1996), 'Identification of Causal Effects Using Instrumental Variables.', *Journal of the American Statistical Association* **91**(434).

Angrist, J. & Krueger, A. (1999), *Handbook of Labour Economics*, Vol. 3, Amsterdam: Elsevier Science, chapter Empirical Strategies in Labor Economics.

ANRF (2006), 'Municipalities with 100% smokefree laws', *Online Reference, Accessed on Nov. 15, 2006* .

Athey, S. & Imbens, G. (2006), 'Identification and inference in nonlinear difference-in-differences models', *Econometrica* pp. 431–497.

Auld, M. & Grootendorst, P. (2004), 'An empirical analysis of milk addiction', *Journal of Health Economics* **23**(6), 1117–1133.

Baggerly, K. A. (2003), 'Empirical likelihood as a goodness-of-fit measure', *Biometrika* **85**(3), 535–547.

Baimbridge, M. & Darcy, D. (1999), 'MPs' Pay 1911 - 1996: Myths and Realities', *Politics* **19**(2), 71–80.

Barro, R. (1973), 'The control of politicians: An economic model', *Public Choice* **14**(1), 19–42.

Barro, R. & Lee, J. (2001), 'International data on educational attainment: updates and implications', *Oxford Economic Papers* **53**(3), 541–563.

Barrow, R. J. & Lee, J. (1994), 'Data set for a panel of 138 countries', *Harvard University* .

Bartels, L. (1993), 'Messages received: The political impact of media exposure', *American Political Science Review* pp. 267–285.

Bauer, T. K., Lofstrom, M. & Zimmerman, K. F. (2000), 'Immigration policy, assimilation of immigrants, and natives sentiments towards immigrants: Evidence from 12 oecd-countries', *IZA Discussion Paper No. 187* .

Beer, S. (1956), 'Pressure Groups and Parties in Britain', *American Political Science Review* **50**(1), 1–23.

Beer, S. (1965), *Modern British politics*, Faber and Faber London.

Begay, M., Traynor, M. & Glantz, S. (1993), 'The tobacco industry, state politics, and tobacco education in California', *American Journal of Public Health* **83**(9), 1214–1221.

Bennett, W. L. (1998), *Communicating democracy: The media and political transitions*, Lynne Rienner, chapter The media and democratic development: The social basis of political communication.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004), 'How Much Should We Trust Differences-in-Differences Estimates?', *The Quarterly Journal of Economics* **119**(1), 249–275.

Bertrand, M. & Mullainathan, S. (2001), 'Do People Mean What They Say? Implications for Subjective Survey Data', *The American Economic Review* **91**(2), 67–72.

Besley, T. (2005), 'Political Selection', *Journal of Economic Perspectives* **19**(3), 43–60.

Besley, T. (2006), *Principled Agents?: The Political Economy of Good Government*, Oxford University Press, USA.

Besley, T. & Burgess, R. (2002), 'The Political Economy of Government Responsiveness: Theory and Evidence from India*', *Quarterly Journal of Economics* **117**(4), 1415–1451.

Besley, T. & Case, A. (1995), 'Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits', *The Quarterly Journal of Economics* **110**(3), 769–798.

Besley, T. & Coate, S. (1997), 'An Economic Model of Representative Democracy', *Quarterly Journal of Economics* **112**(1).

Bhagwati, J. (2002), *The Wind of the Hundred Days: How Washington Mismanaged Globalization*, MIT Press.

Bikhchandani, S., Hirshleifer, D. & Welch, I. (1998), 'Learning from the behavior of others: Conformity, fads, and informational cascades', *The Journal of Economic Perspectives* pp. 151–170.

Black, S. (1999), 'Do Better Schools Matter? Parental Valuation of Elementary Education*', *Quarterly Journal of Economics* **114**(2), 577–599.

Borio, G. (2005), 'Tobacco timeline: Chapter eight, the twenty-first century, the new millenium', *Unpublished Manuscript* .

Borjas, G. (1997), 'How Much Do Immigration and Trade Affect Labor Market Outcomes?', *Brookings Papers on Economic Activity* .

Borjas, G. (1999), *Heavens Door*, Princeton: Princeton University Press.

Borjas, G. (2003), 'The Labor Demand Curve Is Downward Sloping: Reexamining The Impact Of Immigration On The Labor Market', *The Quarterly Journal of Economics* **118**(4), 1335–1374.

Borjas, G. (2005), *Native Internal Migration and the Labor Market Impact of Immigration*, National Bureau of Economic Research Cambridge, Mass., USA.

Borjas, G., Freeman, R. & Katz, L. (1996), 'Searching for the Effect of Immigration on the Labor Market', *The American Economic Review* **86**(2), 246–251.

Braumann, C. (1994), 'Fernsehforschung zwischen Parteilichkeit und Objektivität. Zur Zuschauerforschung in der ehemaligen DDR', *Rundfunk und Fernsehen* **42**(4), 524–541.

Breiman, L. (2001), 'Random Forests', *Machine Learning* **45**(1), 5–32.

Breslow, L. & Johnson, M. (1993), 'California's Proposition 99 on tobacco, and its impact', *Annual Review of Public Health* **14**(1), 585–604.

Brezis, E. & Krugman, P. (1993), 'Immigration, Investment, and Real Wages', *NBER Working Paper* **4563**.

Brinks, D. & Coppedge, M. (2006), 'Diffusion is no illusion: Neighbor emulation in the third wave of democracy', *Comparative Political Studies* **39**(4), 463.

Brookhart, M., Schneeweiss, S., Rothman, K., Glynn, R., Avorn, J. & Sturmer, T. (2006), 'Variable Selection for Propensity Score Models', *American Journal of Epidemiology* **163**(12), 1149–1156.

Broszat, M. & Weber, H. (1993), *SBZ-Handbuch: Staatliche Verwaltungen, Parteien, gesellschaftliche Organisationen und ihre Führungskräfte in der Sowjetischen Besatzungszone Deutschlands 1945–1949*, Vol. 2, Oldenbourg.

Brown, A. (2004), *Ending the Cold War*, Palgrave Macmillan, chapter Gorbachev and the end of the Cold War.

Buhl, D. (1990), 'Window to the west: How television from the federal republic influenced events in east germany', *Discussion paper D-5, Joan Shorenstein Barone Center, Harvard University.* .

Burns, P. & Gimpel, J. (2000), 'Economic insecurity, prejudicial stereotypes, and public opinion on immigration policy', *Political Science Quarterly* **115**(3), 201–225.

California Department of Health Services (2006), 'Fast facts', *Online Reference* .

Capehart, T. (2001), 'Trends in the cigarette industry after the master settlement agreement', *USDA Electronic Outlook Report* .

Card, D. (1990), 'The Impact of the Mariel Boatlift on the Miami Labor Market', *Industrial and Labor Relations Review* **43**(2), 245–257.

Card, D. (2001), 'Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration', *Journal of Labor Economics* **19**(1), 22–64.

Card, D. (2005), 'Is the New Immigration Really So Bad?', *The Economic Journal* **115**(507), F300–F323.

Card, D. (2007), How Immigration Affects US Cities, Technical report, CReAM Discussion Paper.

Caselli, F. & Morelli, M. (2004), 'Bad politicians', *Journal of Public Economics* **88**(3-4), 759–782.

Chandler, C. R. & Tsai, Y.-M. T. (2001), 'Social factors influencing immigration attitudes: an analysis of data from the general social survey', *Social Science Journal* **38**(2), 177–188.

Choi, J. & Thum, M. (2007), 'The Economics OF Politically Connected Firms', *Presented at CESifo Area Conference on Public Sector Economics* .

Citrin, J., Green, D. P., Muste, C. & Wong, C. (1997), 'Public opinion towards immigration reform: the role of economic motivations', *Journal of Politics* **59**, 858–881.

Conley, T. & Taber, C. (2008), 'Inference with difference in differences with a small number of policy changes', *Mimeo, University of Chicago* .

Corrales, J. & Westhoff, F. (2006), 'Information technology adoption and political regimes', *International Studies Quarterly* **50**(4), 911–933.

Courtney, A. (1968), *Sailor in a Russian frame*, London: Johnson.

Dal Bó, E., Dal Bó, P. & Di Tella, R. (2006), 'Plata o Plomo?: Bribe and Punishment in a Theory of Political Influence', *American Political Science Review* **100**(01), 41–53.

Dalton, R. (1994), 'Communists and democrats: Democratic attitudes in the two Germanies', *British Journal of Political Science* pp. 469–493.

Darschin, W. & Zubayr, C. (2000), 'Warum sehen die Ostdeutschen anders fern als die Westdeutschen?', *Media Perspektiven* **6**, 249–257.

Davison, A. & Hinkley, D. (1997), *Bootstrap methods and their application*, Cambridge Univ Pr.

Dehejia, R. (2005), 'Practical propensity score matching: a reply to Smith and Todd', *Journal of Econometrics* **125**(1-2), 355364.

Dehejia, R. H. & Wahba, S. (2002), 'Propensity score matching methods for nonexperimental causal studies', *Review of Economics and Statistics* **84**(1), 151–161.

Dehejia, R. & Wahba, S. (1997), *Econometric Methods for Program Evaluation*, Ph. D. Dissertation, Harvard University, chapter Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs.

Dehejia, R. & Wahba, S. (1999), 'Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs', *Journal of the American Statistical Association* **94**, 1053–1062.

DellaVigna, S. & Kaplan, E. (2007), 'The Fox News Effect: Media Bias and Voting*', *The Quarterly Journal of Economics* **122**(3), 1187–1234.

Deming, W. & Stephan, F. (1940), 'On the least squares adjustment of a sampled frequency table when the expected marginal totals are known', *Ann. Math. Statist.* **1940**, 427–444.

Diamond, A. J. & Sekhon, J. (2006), Genetic matching for causal effects: A general multivariate matching method for achieving balance in observational studies. Unpublished manuscript, Dept. of Political Science, UC Berkeley.

Diamond, L. (1993*a*), *Global transformation and the Third World*, Lynne Rienner, chapter The globalization of democracy.

Diamond, L. (1993*b*), *Political culture and democracy in developing*, Lynne Rienner.

Diermeier, D., Keane, M. & Merlo, A. (2005), 'A Political Economy Model of Congressional Careers', *American Economic Review* **95**(1), 347–373.

Dinan, J. & Heckelman, J. (2005), 'The anti-tobacco movement in the Progressive Era: A case study of direct democracy in Oregon', *Explorations in Economic History* **42**(4), 529–546.

DiNardo, J. & Pischke, J. (1997), 'The returns to computer use revisited: Have pencils changed the wage structure too?', *The Quarterly Journal of Economics* pp. 291–303.

Doig, A. (1984), *Corruption and Misconduct in Contemporary British politics*, Penguin Books.

Donald, S. & Lang, K. (2007), 'Inference with difference-in-differences and other panel data', *The Review of Economics and Statistics* **89**(2), 221–233.

Doorenspleet, R. (2004), 'The structural context of recent transitions to democracy', *European Journal of Political Research* **43**(3), 309–335.

Drake, C. (1993), 'Effects of Misspecification of the Propensity Score on Estimators of Treatment Effect', *Biometrics* **49**(4), 1231–1236.

Dunning, T. (2008), 'Improving causal inference: Strengths and limitations of natural experiments', *Political Research Quarterly* **61**(2), 282.

Dustmann, C. & Preston, I. P. (2006), 'Is immigration good or bad for the economy? analysis of attitudinal responses', *Research in Labor Economics* **24**, 3–34.

Dustmann, C. & Preston, I. P. (2007), 'Racial and economic factors in attitudes to immigration', *The B.E. Journal of Economic Analysis & Policy* **7**(1), Article 62.

Eckstein, H. (1975), *Handbook of political science*, Vol. 7, Addison-Wesley, chapter Case studies in political science.

Eisenfeld, B., Kowalczuk, I.-S. & Neubert, E. (2004), *Die verdrängte Revolution: der Platz des 17. Juni 1953 in der deutschen Geschichte*, Bremen: Edition Temmen.

Eliason, M. & Storrie, D. (2006), 'Lasting or latent scars? swedish evidence on the long-term effects of job displacement.', *Journal of Labor Economics* **24**(4), 831 – 856.

Elliott, M., Beckett, M., Chong, K., Hambarsoomians, K. & Hays, R. (2008), 'How Do Proxy Responses and Proxy-Assisted Responses Differ from What Medicare Beneficiaries Might Have Reported about Their Health Care?', *Health Services Research* **43**(3), 833.

Entman, R. & MyiLibrary (1989), *Democracy without citizens: Media and the decay of American politics*, Oxford University Press.

Eriksson, J. (1980), 'A note on solution of large sparse maximum entropy problems with linear equality constraints', *Mathematical Programming* **18**(1), 146–154.

Erlander, S. (1977), 'Entropy in linear programs–an approach to planning', *Report LiTH-MAT-R-77-3, Department of Mathematics, Linkrping University* .

Espenshade, T. J. & Hempstead, K. (1996), 'Contemporary american attitudes toward u.s. immigration', *International Migration Review* **30**, 535–570.

Evans, W., Farrelly, M. & Montgomery, E. (1999), 'Do workplace smoking bans reduce smoking?', *American Economic Review* pp. 728–747.

Facchini, G. & Mayda, A. (2007), 'Individual attitudes towards immigrants: Welfare-state determinants across countries', *Review of Economics and Statistics* **Forthcoming**.

Faccio, M. (2006), 'Politically Connected Firms', *The American economic review* **96**(1), 369–386.

Farrelly, M., Evans, W. & Sfekas, A. (1999), 'The impact of workplace smoking bans: results from a national survey'.

Feenberg, D. & Coutts, E. (1993), 'An introduction to the taxsim model", *Journal of Policy Analysis and Management* **12**(1).

Ferguson, T. & Voth, H. (2008), 'Betting on Hitler-The Value of Political Connections in Nazi Germany*', *Quarterly Journal of Economics* **123**(1), 101–137.

Ferraz, C. & Finan, F. (2008), 'Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance', *MIMEO* .

Fetzer, J. S. (2000), *Public Attitudes Toward Immigration in the United States, France, and Germany*, Cambridge: Cambridge University Press.

Fichtenberg, C. & Glantz, S. (2000), 'Association of the California Tobacco Control Program with declines in cigarette consumption and mortality from heart disease', *New England Journal of Medicine* **343**(24), 1772–1777.

Finer, S. (1962), *Anonymous Empire: A Study of the Lobby in Great Britain*, Pall Mall Press.

Fiorina, M. (1994), 'Divided Government in the American States: A Byproduct of Legislative Professionalism?', *The American Political Science Review* **88**(2), 304–316.

Fisher, J. (1994), 'Political Donations to the Conservative Party', *Parliamentary Affairs* **47**(1), 61–72.

Fisher, R. A. (1925), *The Design of Experiments*, 1st edn, Oliver and Boyd, London.

Fisman, R. (2001), 'Estimating the Value of Political Connections', *American Economic Review* **91**(4), 1095–1102.

Fix, M. & Passel, J. (2002), 'The scope and impact of welfare reforms immigrant provisions', *Discussion Paper 02-03. Washington DC: Urban Institute* .

Fix, M., Passel, J. & Enchautegui, M. (1994), *Immigration and immigrants: setting the record straight*, Urban Institute.

Fletcher, R. (1987), *Practical methods of optimization*, Wiley-Interscience New York, NY, USA.

Frank, K., Sykes, G., Anagnostopoulos, D., Cannata, M., Chard, L., Krause, A. & McCrory, R. (2008), 'Does NBPTS Certification Affect the Number of Colleagues a Teacher Helps With Instructional Matters?', *Educational Evaluation and Policy Analysis* **30**(1), 3.

Freedman, D. (2005), *Statistical models: Theory and practice*, Cambridge University Press.

Fricke, K. W. (1991), *MfS Intern: Macht, Strukturen, Auflösung der DDR-Staatssicherheit*, Köln: Wissenschaft und Politik.

Fricke, K. W. & Engelmann, R. (2003), *Der "Tag X" und die Staatssicherheit: 17. Juni 1953 — Reaktionen und Konsequenzen im DDR-Machtapparat*, Bremen: Edition Temmen.

Friedberg, R. & Hunt, J. (1995), 'The Impact of Immigrants on Host Country Wages, Employment and Growth', *The Journal of Economic Perspectives* **9**(2), 23–44.

Friedrich, W. (1990), 'Mentalitätswandlungen der jugend in der ddr', *Aus Politik und Zeitgeschichte. Beilage zur Wochenzeitung Das Parlament* **16**(7), 25–37.

Friedrich, Walter, P. F. & Starke, K., eds (1999), *Das Zentralinstitut für Jugendforschung Leipzig 1966–1990.*

Frölich, M. (2004), 'Finite Sample Properties of Propensity-Score Matching and Weighting Estimators', *Review of Economics and Statistics* **86**(1), 77–90.

Frölich, M. (2007), 'Propensity score matching without conditional independence assumptionwith an application to the gender wage gap in the United Kingdom', *The Econometrics Journal* **10**(2), 359–407.

Früh, W. & Stiehler, H.-J. (2002), *Fernsehen in Ostdeutschland*, Vistas.

Gabel, M. & Scheve, K. (2007), 'Estimating the Effect of Elite Communications on Public Opinion Using Instrumental Variables', *American Journal of Political Science* pp. 1013–1028.

Gabriel, O. & Troitzsch, K. (1993), *Wahlen in Zeiten des Umbruchs*, Lang.

Gagliarducci, S., Nannicini, T. & Naticchioni, P. (2008), 'Outside Income and Moral Hazard: The Elusive Quest for Good Politicians'.

Gang, I. N., Rivera-Batiz, F. L. & Yun, M.-S. (2002), 'Economic strain, ethnic concentration and attitudes towards foreigners in the european union.', *IZA Discussion Paper No. 578* .

Gasiorowski, M. (1995), 'Economic crisis and political regime change: An event history analysis', *American Political Science Review* pp. 882–897.

Geddes, B. & Zaller, J. (1989), 'Sources of popular support for authoritarian regimes', *American Journal of Political Science* pp. 319–347.

Gentzkow, M. (2006), 'Television and Voter Turnout', *The Quarterly Journal of Economics* **121**(3), 931–972.

George, A. & Bennett, A. (2005), *Case studies and theory development in the social sciences*, Mit Press.

Giddens, A. (2003), *Runaway world: how globalisation is reshaping our lives*, Routledge.

Glantz, S. (1993), 'Changes in cigarette consumption, prices, and tobacco industry revenues associated with California's Proposition 99', *British Medical Journal* **2**(4), 311–314.

Glantz, S. & Balbach, E. (2000), *Tobacco war: inside the California battles*, University of California Press.

Gleditsch, K. & Ward, M. (2006), 'Diffusion and the international context of democratization', *International Organization* **60**(04), 911–933.

Goel, R. & Nelson, M. (2006), 'The effectiveness of anti-smoking legislation: a review', *Journal of Economic Surveys* **20**(3), 325–355.

Goldman, E., Rocholl, J. & So, J. (2008*a*), 'Does political connectedness affect firm value?', *Review of Financial Studies* **forthcoming**.

Goldman, E., Rocholl, J. & So, J. (2008*b*), 'Political Connections and the Allocation of Procurement Contracts', *Working Paper University of Indiana* .

Granovetter, M. (1978), 'Threshold models of collective behavior', *American journal of sociology* **83**(6), 1420.

Gray, C. (1999), 'Clausewitz rules, OK? The future is the pastwith GPS', *Review of International Studies* **25**(05), 161–182.

Groseclose, T. & Krehbiel, K. (1994), 'Golden Parachutes, Rubber Checks, and Strategic Retirements from the 102d House', *American Journal of Political Science* **38**(1), 75–99.

Grossman, G. & Helpman, E. (1994), 'Protection for Sale', *American Economic Review* **84**(4), 833–850.

Gruber, J. (2001), 'Tobacco at the crossroads: The past and future of smoking regulation in the United States', *Journal of Economic Perspectives* pp. 193–212.

Grundmann, S. (1997), *Regionale Strukturen im Wandel*, Opladen: Leske und Budrich, chapter Territorialplanung in der DDR: Indikatoren zur Analyse regionaler Disparitäten — Die sozial-räumliche Struktur der DDR in den 80er Jahren.

Grundmann, S. (1998), *Bevölkerungsentwicklung in Ostdeutschland*, Opladen: Leske und Budrich.

Grünert, H. (1996), *Arbeit, Arbeitsmarkt und Betriebe*, Opladen: Leske und Budrich, chapter Beschäftigungssystem der DDR.

Grünert, H., Bernien, M. & Lutz, B. (1997), *Der ostdeutsche Arbeitsmarkt in Gesamtdeutschland: Angleichung oder Auseinanderdriften?*, chapter Beschäftigungssystem der DDR: Funktionsweise, Entwicklungstendenzen und Folgewirkungen.

Gu, X. & Rosenbaum, P. (1993), 'Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms', *Journal of Computational and Graphical Statistics* **2**(4), 405–420.

Gunther, R. & Mughan, A. (2000), *Democracy and the media: a comparative perspective*, Cambridge Univ Pr.

Hahn, J. (1998), 'On the role of the propensity score in efficient semiparametric estimation of average treatment effects', *Econometrica* **66**, 315331.

Hainmueller, J. & Hiscox, M. (2006), 'Learning to Love Globalization: Education and Individual Attitudes Toward International Trade', *International Organization* **60**(02), 469–498.

Hainmueller, J. & Hiscox, M. J. (2007), 'Educated preferences: Explaining attitudes toward immigration in europe', *International Organization* **61**(2), 399–442.

Hainmueller, J. & Kern, H. (2008), 'Incumbency as a source of spillover effects in mixed electoral systems: Evidence from a regression-discontinuity design', *Electoral Studies* **27**, 213–227.

Hajna, K. (2000), *Die Landtagswahlen 1946 in der SBZ*, Lang.

Hall, R. & van Houweling, R. (1995), 'Avarice and Ambition in Congress: Representatives' Decisions to Run or Retire from the US House', *The American Political Science Review* **89**(1), 121–136.

Hansen, L. (1982), 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica* **50**(4), 1029–1054.

Hanson, G. H. (2005), *Why Does Immigration Divide America?* ., Washington DC: Institute for International Economics.

Hanson, G., Scheve, K. & Slaughter, M. (2007), 'Public finance and individual preferences over globalization strategies', *Economics and Politics* **19**(1), 1–33.

Hanson, G., Scheve, K. & Slaughter, M. (2008), *Skilled Migration Today: Prospects, Problems, and Policies*, New York: Council on Foreign Relations Press, chapter Individual Preferences over High-Skilled Immigration in the United States.

Harrison, M. (1960), *Trade Unions and the Labour Party Since 1945*, Allen & Unwin.

Harwood, E. (1986), 'American public opinion and u.s. immigration policy', *Annals of the American Academy of Political and Social Science* **487**, 201–212.

Heckman, J. & Hotz, V. (1989), 'Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training', *Journal of the American Statistical Association* pp. 862–874.

Hellerstein, J. & Imbens, G. (1999), 'Imposing Moment Restrictions from Auxiliary Data by Weighting', *The Review of Economics and Statistics* **81**(1), 1–14.

Hero, R. & Preuhs, R. (2007), 'Immigration and the Evolving American Welfare State: Examining Policies in the US States', *American Journal of Political Science* **51**(3), 498–517.

Hesse, K. (1988), *Westmedien in der DDR: Nutzung, Image und Auswirkungen bundesrepublikanischen Hörfunks und Fernsehens*, Verlag Wissenschaft und Politik.

Hesse, K. (1990), 'Cross-border mass communication from West to East Germany', *European Journal of Communication* **5**(2), 355.

Heym, S. (1977), 'Je voller der mund, desto leerer die sprüche', *Stern* **February 10**, 104–110.

Hirano, K. & Imbens, G. (2001), 'Estimation of causal effects using propensity score weighting: An application of data on right hear catherization', *Health Services and Outcomes Research Methodology* **2**, 259–278.

Hirano, K., Imbens, G. & Ridder, G. (2003), 'Efficient estimation of average treatment effects using the estimated propensity score', *Econometrica* **71**(4), 1161–1189.

HMRC (2007), Quality Review of Personal Wealth National Statistics, Technical report, HM Revenue & Customs.

Ho, D., Imai, K., King, G. & Stuart, E. (2007), 'Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference', *Political Analysis* **15**(3), 199.

Holland, P. W. (1986), 'Statistics and Causal Inference (with discussion)', *Journal of the American Statistical Association* **81**(396), 945–960.

Hollingsworth, M. (1991), *MPs for Hire: The Secret World of Political Lobbying*, Bloomsbury.

Horvitz, D. & Thompson, D. (1952), 'A Generalization of Sampling Without Replacement From a Finite Universe', *Journal of the American Statistical Association* **47**(260), 663–685.

Hu, T., Sung, H. & Keeler, T. (1995), 'Reducing cigarette consumption in California: tobacco taxes vs an anti-smoking media campaign'.

Huntington, S. (1991), *The third wave: Democratization in the late twentieth century*, University of Oklahoma Press.

Iacus, S., King, G. & Porro, G. (2008), 'Matching for causal inference without balance checking', *Mimeo Harvard University* .

Ibragimov, R. & Sharakhmetov, S. (1998), 'On an exact constant for the Rosenthal inequality', *Theory of Probability and its Applications* **42**(2), 294–301.

Imai, K., King, G. & Stuart, E. (2008), 'Misunderstandings among experimentalists and observationalists: Balance test fallacies in causal inference', *J. R. Statist. Soc. A* **171**, 481–502.

Imbens, G. (1997), 'One-Step Estimators for Over-Identified Generalized Method of Moments Models', *The Review of Economic Studies* **64**(3), 359–383.

Imbens, G. (2004), 'Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review', *Review of Economics and Statistics* **86**(1), 4–29.

Imbens, G. & Angrist, J. (1994), 'Identification and estimation of local average treatment effects', *Econometrica: Journal of the Econometric Society* pp. 467–475.

Imbens, G. & Lemieux, T. (2007), 'Regression Discontinuity Designs: A Guide to Practice', *NBER Working Paper* **13039**.

Imbens, G. M. & Wooldridge, J. M. (2008), Recent developments in the econometrics of program evaluation, Working Paper 14251, National Bureau of Economic Research.

Imbens, G., Spady, R. & Johnson, P. (1998), 'Information Theoretic Approaches to Inference in Moment Condition Models', *Econometrica* **66**(2), 333–357.

Indicators, D. & Analyses, S. (2007), 'Quality of Diabetes Care Among Cancer Survivors With Diabetes', *Medical Care* **45**(9), 869–875.

Ireland, C. & Kullback, S. (1968), 'Contingency tables with given marginals', *Biometrika* **55**, 179–188.

Jarausch, K. (1994), *The rush to German unity*, Oxford University Press, USA.

Jayachandran, S. (2006), 'The Jeffords Effect', *The Journal of Law and Economics* **49**(2), 397–425.

Johnson, S. & Mitton, T. (2003), 'Cronyism and capital controls: evidence from Malaysia', *Journal of Financial Economics* **67**(2), 351–382.

Jones, R. (1971), 'A Three-Factor Model in Theory, Trade, and History', *Trade, Balance of Payments, and Growth* pp. 3–21.

Judge, D. (1984), 'The Politics of MPs' Pay', *Parliamentary Affairs* **37**(1), 59–75.

Kalathil, S. & Boas, T. (2003), *Open networks, closed regimes: The impact of the Internet on authoritarian rule*, Carnegie Endowment.

Kapur, J. & Kevsavan, H. (1992), *Entropy optimization principles with applications*, London: Academic Press.

Katz, E. & Foulkes, D. (1962), 'On the use of the mass media as" escape": Clarification of a concept', *Public Opinion Quarterly* **26**(3), 377–388.

Keane, M. & Merlo, A. (2007), 'Money, Political Ambition, and the Career Decisions of Politicians', *Mimeo University of Pennsylvania* .

Kessler, A. (2001), 'Immigration, economic insecurity, and the "ambivalent" american public', *Center for Comparative Immigration Studies. Working Paper.* .

Khwaja, A. & Mian, A. (2005), 'Do Lenders Favor Politically Connected Firms? Rent Provision in an Emerging Financial Market*', *The Quarterly Journal of Economics* **120**(4), 1371–1411.

King, G., Murray, C., Salomon, J. & Tandon, A. (2004), 'Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research', *American Political Science Review* **98**(01), 191–207.

King, G. & Zeng, L. (2006), 'The dangers of extreme counterfactuals', *Political Analysis* **14**(2), 131–159.

Kitamura, Y. & Stutzer, M. (1997), 'An Information-Theoretic Alternative to Generalized Method of Moments Estimation', *Econometrica* **65**(4), 861–874.

Knabe, H. (2003), *17. Juni 1953: ein deutscher Aufstand*, Propyläen.

Kopstein, J. & Reilly, D. (2000), 'Geographic diffusion and the transformation of the postcommunist world', *World Politics* pp. 1–37.

Kowalczuk, I., Mitter, A. & Wolle, S. (1996), *Der Tag X, 17. Juni 1953: Die innere Staatsgründung" der DDR als Ergebnis der Krise 1952/54*, Ch. Links.

Kroszner, R. & Strahan, P. (1999), 'What Drives Deregulation? Economics and Politics of The Relaxation Of Bank Branching Restrictions', *Quarterly Journal of Economics* **114**(4), 1437–1467.

Krugman, P. & Obstfeld, M. (2000), *International economics: theory and policy*, 5 edn, Reading, MA: Addison-Wesley Longman.

Kullback, S. (1959), *Information Theory and Statistics*, Wiley, NY.

Kuran, T. (1991), 'Now out of never: The element of surprise in the East European revolution of 1989', *World Politics: A Quarterly Journal of International Relations* pp. 7–48.

LaLonde, R. J. (1986), 'Evaluating the Econometric Evaluations of Training Programs with Experimental Data', *American Economic Review* **76**, 604–620.

Lazarsfeld, P. F. & Merton, R. K. (1948), *Communication of ideas*, Harper & Brothers, chapter Mass communication, popular taste and organized social action.

Leamer, E. & Levinsohn, J. (1995), *International Trade Theory: The Evidence. In Handbook of International Economics*, Vol. 3, Amsterdam: North-Holland, chapter International Trade Theory: The Evidence, pp. 1339–1394.

Lee, D. (2008), 'Randomized experiments from non-random selection in US House elections', *Journal of Econometrics* **142**(2), 675–697.

Lee, D. S., Moretti, E. & Butler, M. J. (2004), 'Do Voters Affect or Elect Policies? Evidence from the U.S. House', *Quarterly Journal of Economics* **119**(3), 807–859.

Lehmann, E. (1997), *Testing Statistical Hypotheses*, Vol. 2, Berkeley: University of California Press.

Lévesque, J. (2004), *Ending the Cold War*, Palgrave Macmillan, chapter The emancipation of Eastern Europe.

Lewis, E. (2005), *Immigration, Skill Mix, and the Choice of Technique*, Federal Reserve Bank of Philadelphia, Research Dept.

Lisann, M. (1975), *Broadcasting to the Soviet Union: international politics and radio*, Praeger Publishers.

Lohmann, S. (1994), 'The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989-91', *World Politics* pp. 42–101.

Longhi, S., Nijkamp, P. & Poot, J. (2005), 'A Meta-Analytic Assessment of the Effect of Immigration on Wages', *Journal of Economic Surveys* **19**(3), 451–477.

Longo, D., Johnson, J., Kruse, R., Brownson, R. & Hewett, J. (2001), 'A prospective investigation of the impact of smoking bans on tobacco cessation and relapse', *British Medical Journal* **10**(3), 267–272.

Mattos, R. & Veiga, A. (2004), 'Entropy optimization: Computer implementation of the maxent and minexent principles', *Unpublished Manuscript* .

Mattozzi, A. & Merlo, A. (2007), *Political Careers Or Career Politicians?*, National Bureau of Economic Research Cambridge, Mass., USA.

Mayda, A. (2006), 'Who is against immigration? a cross-country investigation of individual attitudes toward immigrants', *Review of Economics and Statistics* **88**(3), 510–530.

Mayda, A. & Rodrik, D. (2005), 'Why are some people (and countries) more protectionist than others?', *European Economic Review* **49**(6), 1393–1430.

Mayer, W. (2002), *Flucht und Ausreise: Botschaftsbesetzungen als Form des Widerstands gegen die Politische Verfolgung in der DDR*, Berlin: Anita Tykve Verlag.

McCaffrey, D., Ridgeway, G. & Morral, A. (2004), 'Propensity score estimation with boosted regression for evaluating adolescent substance abuse treatment', *Psychological Methods* **9**(4), 403–425.

McNiel, D. & Binder, R. (2007), 'Effectiveness of a Mental Health Court in Reducing Criminal Recidivism and Violence', *American Journal of Psychiatry* **164**(9), 1395.

Meinhardt, V., Seidel, B., Stile, F. & Teichmann, D. (1995), 'Transferleistungen in die neuen bundesländer und deren wirtschaftliche konsequenzen', *Deutsches Institut für Wirtschaftfroschung, Sonderheft 154* .

Menchik, P. & David, M. (1983), 'Income Distribution, Lifetime Savings, and Bequests', *The American Economic Review* **73**(4), 672–690.

Merlo, A. (2006), 'Whither Political Economy? Theories, Facts and Issues', *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, Cambridge: Cambridge University Press, forthcoming* .

Messner, M. & Polborn, M. (2004), 'Paying Politicians', *Journal of Public Economics* **88**(12), 2423–2445.

Meyen, M. (2001), 'Haben die Westmedien die DDR stabilisiert?', *Siegener Periodicum zur internationalen empirischen Literaturwissenschaft* **20**(1), 117–133.

Meyen, M. (2003*a*), *Denver Clan und Neues Deutschland.*

Meyen, M. (2003*b*), *Einschalten, umschalten, ausschalten?: Das Fernsehen im DDR-alltag*, Leipziger Universitätsverlag.

Milman, M., Jiang, F. & Jelliffe, R. (2001), 'Creating discrete joint densities from continuous ones: the moment matching-maximum entropy approach', *Computers in Biology and Medicine* **31**(3), 197–214.

Mondak, J. (1995), 'Media exposure and political discussion in US elections', *Journal of Politics* pp. 62–85.

Morgan, S. & Winship, C. (2007), *Counterfactuals and causal inference: Methods and principles for social research*, Cambridge Univ Pr.

Muller, W. (1977), *The Kept Men?: The First Century of Trade Union Representation in the British House of Commons, 1874-1975*, Harvester Press.

Nelson, M. (1997), *War of the black heavens: the battles of Western broadcasting in the Cold War*, Syracuse Univ Pr.

Newey, W. & Smith, R. (2004), 'Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators', *Econometrica* **72**(1), 219–255.

Newton, K. (1999), 'Mass media effects: mobilization or media malaise?', *British Journal of Political Science* **29**(04), 577–599.

Noel-Baker, P. (1961), 'The Grey Zone: The Problems of Business Affiliations of Members of Parliament', *Parliamentary Affairs, v15* .

Norton, P. (2003), 'The United Kingdom: Restoring Confidence?', *Parliamentary Affairs* **50**(3), 357–372.

Nye, J. (2006), *Soft power: the means to success in world politics*, Perseus Books Group.

Nye Jr, J. (2008), 'Public Diplomacy and Soft Power', *The annals of the American academy of political and social science* **616**(1), 94.

O'Donnell, G., Schmitter, P. & Whitehead, L. (1986), *Transitions from authoritarian rule: prospects for democracy*, Johns Hopkins University Press.

Oh, H. L. & Scheuren, F. J. (1978), 'Multivariate ratio raking estimation in the 1973 exact match study', *Proceedings of the Section on Survey Research Methods, American Statistical Association* pp. 716–722.

Olken, B. (2008), 'Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages', *NBER Working Paper* .

O'Loughlin, J., Ward, M., Lofdahl, C., Cohen, J., Brown, D., Reilly, D., Gleditsch, K. & Shin, M. (1998), 'The diffusion of democracy, 1946–1994', *Annals of the Association of American Geographers* **88**(4), 545–574.

O'Neil, P. H. (1998), *Communicating democracy: The media and political transitions*, Lynne Rienner, chapter Democratization and mass communication: What is the link?

Opp, K., Voß, P. & Gern, C. (1993), *Die volkseigene Revolution*, Klett-Cotta.

Orzechowski & Walker (2005), 'The tax burden on tobacco. historical compilation', *Arlington, VA: Orzechowski & Walker.* **40**.

Osborne, M. & Slivinski, A. (1996), 'A Model of Political Competition with Citizen-Candidates', *The Quarterly Journal of Economics* **111**(1), 65–96.

Ottaviano, G. & Peri, G. (2008), 'Immigration and National Wages: Clarifying the Theory and the Empirics', *NBER WP 14188* .

Owen, A. (2001), *Empirical Likelihood*, Chapman & Hall/CRC.

Owens, A., Green, D., Bailey, C. & Kay, A. (2006), 'A measure of worth: probate valuations, personal wealth and indebtedness in England, 1810-40', *Historical Research* **79**(205), 382–403.

Parta, R. E. (2007), *Discovering the hidden listener: An assessment of Radio Liberty and Western Broadcasting to the USSR during the Cold War*, Hoover Institution Press.

Patterson, T. (1994), *Out of order*, Vintage Books New York.

Peltzman, S. (1984), 'Constituent Interest and Congressional Voting', *The Journal of Law and Economics* **27**(1), 181.

Pfaff, S. (2006), *Exit-voice dynamics and the collapse of East Germany: the crisis of Leninism and the revolution of 1989*, Duke Univ Pr.

Pfaff, S. & Kim, H. (2003), 'Exit-Voice Dynamics in Collective Action: An Analysis of Emigration and Protest in the East German Revolution 1', *American Journal of Sociology* **109**(2), 401–444.

Pierce, J., Gilpin, E., Emery, S., Farkas, A., Zhu, S., Choi, W., Berry, C., Distefan, J., White, M., Soroko, S. et al. (1998), 'Tobacco control in California: Whos winning the war? An evaluation of the Tobacco Control Program, 1989–1996', *La Jolla, CA: University of California, San Diego* pp. 12–18.

Pinto-Duschinsky, M. (1981), *British Political Finance, 1830-1980*, American Enterprise Institute for Public Policy Research.

Pinto-Duschinsky, M. (1989), 'Trends in British Party Funding 1983-1987', *Parliamentary Affairs* **42**(2), 197–212.

Pinto-Duschinsky, M. (1990), *UK Political Parties Since 1945*, Philip Allan: London, chapter The funding of political parties since 1945, pp. 95–109.

Postman, N. (1986), *Amusing ourselves to death*, Penguin Books New York, NY, USA.

Pridham, G. (1991), *Encouraging democracy: the international context of regime transition in Southern Europe*, Burns & Oates.

Pridham, G. (1997), *Building democracy? The international dimension of democratization in Eastern Europe*, Leicester University Press, chapter The international dimension of democratization: theory, practice, and inter-regional comparisons.

Puddington, A. (2003), *Broadcasting Freedom: The Cold War Triumph of Radio Free Europe and Radio Liberty*, Univ Pr of Kentucky.

Qin, J. & Lawless, J. (1994), 'Empirical Likelihood and General Estimating Equations', *The Annals of Statistics* **22**(1), 300–325.

Querubin, P. & Snyder, J. M. (2008), The Rents to Political Office in the U.S., 1840-1870. Manuscript, Massachusetts Institute of Technology.

Quester, G. (1990), *The international politics of television*, Free Press.

R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Read, T. & Cressie, N. (1988), *Goodness-Of-Fit Statistics for Discrete Multivariate Data*, Springer.

Redelmeier, D. & Singh, S. (2001), 'Longevity of screenwriters who win an academy award: longitudinal study'.

Ridgeway, G. (2006), 'Assessing the Effect of Race Bias in Post-traffic Stop Outcomes Using Propensity Scores', *Journal of Quantitative Criminology* **22**(1), 1–29.

Roberts, A. (1991), *Civil resistance in the East European and Soviet revolutions*, Albert Einstein Institution.

Roberts, B. (1990), 'A Dead Senator Tells No Lies: Seniority and the Distribution of Federal Benefits', *American Journal of Political Science* **34**(1), 31–58.

Robins, J. & Ritov, Y. (1997), 'Toward a Curse of Dimensionality Appropriate(Coda) Asymtotic Theory for Semi-Parametric Models', *Statistics in Medicine* **16**(3), 285–319.

Robins, J. & Rotnitzky, A. (1995), 'Semiparametric Efficiency in Multivariate Regression Models with Missing Data.', *Journal of the American Statistical Association* **90**(429).

Robins, J., Rotnitzky, A. & Zhao, L. (1995), 'Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data.', *Journal of the American Statistical Association* **90**(429).

Rohde, D. (1979), 'Risk-Bearing and Progressive Ambition: The Case of Members of the United States House of Representatives', *American Journal of Political Science* **23**(1), 1–26.

Rohrschneider, R. (1994), 'Report from the laboratory: The influence of institutions on political elites' democratic values in Germany', *American Political Science Review* pp. 927–941.

Rohrschneider, R. (1996), 'Institutional learning versus value diffusion: The evolution of democratic values among parliamentarians in Eastern and Western Germany', *Journal of Politics* pp. 422–446.

Rosenbaum, P. (1984), 'The Consquences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment', *Journal of the Royal Statistical Society. Series A (General)* pp. 656–666.

Rosenbaum, P. (1987), 'Model-based direct adjustment', *Journal of the American Statistical Association* **82**, 387–394.

Rosenbaum, P. (2002*a*), 'Covariance adjustment in randomized experiments and observation', *Statistical Science* **17**(3), 286–327.

Rosenbaum, P. R. (2002*b*), *Observational Studies*, 2nd edn, New York: Springer-Verlag.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika* **70**(1), 41–55.

Rosenbaum, P. & Rubin, D. (1984), 'Reducing Bias in Observational Studies Using Subclassification on the Propensity Score', *Journal of the American Statistical Association* **79**(387), 516–524.

Roth, A. (1957), *The Business Background of MPs*, Parliamentary Profile Services.

Roth, A. (1965), *The Business Background of MPs*, Parliamentary Profile Services.

Rubin, D. (1974), 'Estimating causal effects of treatments in randomized and non-randomized studies', *Journal of Educational Psychology* **66**(5), 688–701.

Rubin, D. (2006), *Matched Sampling for Causal Effects*, Cambridge University Press.

Rubin, D. B. (1976*a*), 'Multivariate Matching Methods that are Equal Percent Bias Reducing, I: Some Examples', *Biometrics* **32**(1), 109–120.

Rubin, D. B. (1976*b*), 'Multivariate Matching Methods that are Equal Percent Bias Reducing, II: Some Examples', *Biometrics* **32**(1), 121–132.

Rubin, D. B. (1978), 'Bayesian Inference for Causal Effects: The Role of Randomization', *Annals of Statistics* **6**(1), 34–58.

Rubin, D. B. (2001), 'Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation', *Health Services & Outcomes Research Methodology* **2**, 169188.

Rubin, D. & Stuart, E. (2006), 'Affinely invariant matching methods with discriminant mixtures of proportional ellipsoidally symmetric distributions', *Annals of Statistics* **34**, 1814–1826.

Rubin, D. & Thomas, N. (1992), 'Affinely Invariant Matching Methods with Ellipsoidal Distributions', *The Annals of Statistics* **20**(2), 1079–1093.

Rubin, D. & Thomas, N. (1996), 'Matching Using Estimated Propensity Scores: Relating Theory to Practice', *Biometrics* **52**(1), 249–264.

Rush, M. (1969), *The Selection of Parliamentary Candidates*, Nelson.

Rustow, D. (1989), 'Democracy: a global revolution', *Foreign Affairs* **69**, 75.

Särndal, C. E. & Lundström, S. (2006), *Estimation in Surveys with Nonresponse*, John Wiley & Sons, Ltd.

Schennach, S. (2007), 'Point estimation with exponentially tilted empirical likelihood', *The Annals of Statistics* **35**(2), 634–672.

Scheve, K. & Slaughter, M. (2001), 'Labor market competition and individual preferences over immigration policy', *Review of Economics and Statistics* **83**(1), 133–145.

Schlesinger, J. (1966), *Ambition and Politics: Political Careers in the United States*, Rand McNally.

Schmitt, K. (1993), *Wahlen in Zeiten des Umbruchs*, Peter Lang, chapter Politische Landschaften im Umbruch: Das Gebiet der ehemaligen DDR 1928–1990.

Sekhon, J. (2006), 'Alternative balance metrics for bias reduction in matching methods for causal inference', *Unpublished manuscript, Dept. of Political Science, UC Berkeley* .

Sekhon, J. (2007), 'Multivariate and propensity score matching software with automated balance optimization: The matching package for r', *Journal of Statistical Software* .

Shirk, S. (2007), *China: fragile superpower*, Oxford University Press, USA.

Siegel, M. (2002), 'The effectiveness of state-level tobacco control interventions: A review of program implementation and behavioral outcomes', *Annual review of public health* **23**(1), 45–71.

Smith, J. P. & Edmonston, B., eds (1997), *The New Americans: Economic, Demographic, and Fiscal Effects of Immigration*, Washington DC: National Academies Press.

Smith, J. & Todd, P. (2001), 'Reconciling conflicting evidence on the performance of propensity-score matching methods', *American Economic Review* **91**(2), 112–118.

Smith, J. & Todd, P. (2005*a*), 'Does matching overcome LaLonde's critique of nonexperimental estimators?', *Journal of Econometrics* **125**(1-2), 305–353.

Smith, J. & Todd, P. (2005*b*), 'Rejoinder', *Journal of Econometrics* **125**(1-2), 365–375.

Sola Pool, I. (1974), *Handbook of communication*, Rand McNally, chapter Communication in totalitarian societies.

Staatliche Zentralverwaltung für Statistik (1989), *Statistisches Jahrbuch 1989*, Berlin: Staatsverlag der Deutschen Demokratischen Republik.

Starr, H. (1991), 'Democratic dominoes: Diffusion approaches to the spread of democracy in the international system', *Journal of Conflict Resolution* pp. 356–381.

Starr, H. & Lindborg, C. (2003), 'Democratic dominoes revisited: The hazards of governmental transitions, 1974-1996', *Journal of Conflict Resolution* **47**(4), 490–501.

Stewart, J. (1958), *British Pressure Groups: Their Role in Relation to the House of Commons*, The Clarendon Press.

Stiehler, H. (2001), 'Leben ohne Westfernsehen', *Studien zur Medienwirkung und Mediennutzung in der Region Dresden der 80er Jahre. Leipzig, Media Studien Bd. 9* .

Stroud, N. (2008), 'Media use and political predispositions: Revisiting the concept of selective exposure', *Political Behavior* **30**(3), 341–366.

Sükösd, M. (2000), *Democracy and the media*, Cambridge University Press, chapter Democratic transformation and the mass media in Hungary: From stalinism to democratic consolidation.

Sylvestre, M., Huszti, E. & Hanley, J. (2006), 'Do Oscar Winners Live Longer than Less Successful Peers? A Reanalysis of the Evidence', *Annals of Internal Medicine* **145**(5), 361.

Tate, C. (2000), *Cigarette wars: the triumph of the little white slaver*, Oxford University Press, USA.

Temple, J. (1999), 'The new growth evidence', *Journal of Economic Literature* **3**, 112–156.

Thistlethwaite, D. & Campbell, D. (1960), 'Regression-discontinuity analysis: An alternative to the ex post facto experiment', *Journal of Educational Psychology* **51**, 309–317.

Uunk, W., Mach, B. & Mayer, K. (2005), 'Job mobility in the former east and West Germany: The effects of state-socialism and labor market composition', *European Sociological Review* **21**(4), 393–408.

van de Walle, D. & Mu, R. (2007), 'Fungibility and the flypaper effect of project aid: Micro-evidence for Vietnam', *Journal of Development Economics* **84**(2), 667–685.

Welzel, C. (2006), 'Democratization as an emancipative process: The neglected role of mass motivations', *European Journal of Political Research* **45**(6), 871–896.

Welzel, C. & Inglehart, R. (2008), 'The role of ordinary people in democratization', *Journal of Democracy* **19**(1), 126.

Whaley, J. (1999), Legislative Ethics: A Comparative Analysis, Technical report, National Democratic Institute for International Affairs.

Whitehead, L. (1996), *The international dimensions of democratization*, Oxford University Press, chapter Three international dimensions of democratization.

Wolle, S. (1998), *Die heile Welt der Diktatur: Alltag und Herrschaft in der DDR 1971-1989*, Ch. Links.

Woodruff, T., Rosbrook, B., Pierce, J. & Glantz, S. (1993), 'Lower levels of cigarette consumption found in smoke-free workplaces in California', *Archives of Internal Medicine* **153**(12), 1485–1493.

Wooldridge, J. (2003), 'Cluster-sample methods in applied econometrics', *American Economic Review* **93**(2), 133–138.

Zaslavsky, A. (1998), 'Representing local reweighting area adjustments by of households', *Survey Methodology* **14**(2), 265–288.

Zelikow, P. & Rice, C. (1995), *Germany unified and Europe transformed: a study in statecraft*, Harvard University Press.

Zhao, Z. (2004), 'Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence', *Review of Economics and Statistics* **86**(1), 91–107.

Zimmerman, W. & Tumlin, K. C. (1999), 'Patchwork policies: State assistance for immigrants under welfare reform', *Washington DC: Urban Institute* .