

Essays on the Political Economy of the American Frontier

by

Jason V Poulos

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Political Science

and the Designated Emphasis

in

Computational and Data Science and Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Sean Gailmard, Chair

Assistant Professor Joshua Blumenstock

Professor Eric Schickler

Professor Jasjeet Sekhon

Spring 2019

The dissertation of Jason V Poulos, titled Essays on the Political Economy of the American Frontier, is approved:

Chair _____

Date _____

University of California, Berkeley

Essays on the Political Economy of the American Frontier

Copyright 2019
by
Jason V Poulos

Abstract

Essays on the Political Economy of the American Frontier

by

Jason V Poulos

Doctor of Philosophy in Political Science

and the Designated Emphasis

in

Computational and Data Science and Engineering

University of California, Berkeley

Professor Sean Gailmard, Chair

Your abstract text here. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam sollicitudin ligula at sapien semper quis consectetur justo consequat. Mauris tristique vehicula tortor pellentesque auctor. Vivamus metus mauris, convallis sit amet mattis non, laoreet non lorem. Pellentesque a tempus lacus. Morbi suscipit porttitor tempor. Nulla facilisi. Morbi nunc erat, imperdiet eget dignissim ac, dictum quis nisl. Aenean viverra elit sit amet nulla ornare viverra. Vivamus fermentum, nunc in dignissim porta, nibh tellus viverra lacus, sed malesuada libero purus et velit. Praesent volutpat leo eu risus rutrum posuere. Etiam cursus ultrices enim. Suspendisse fringilla leo ut ligula dapibus ut consequat justo vehicula. Ut vulputate, justo in condimentum molestie, orci arcu posuere urna, vel laoreet augue magna vel tortor. Fusce ut ante lorem, quis dignissim purus. Nam eget ligula quis sapien scelerisque elementum. Quisque congue tempus ligula, id consectetur mi congue viverra.

To Ossie Bernosky

And exposition? Of go. No upstairs do fingering. Or obstructive, or purposeful. In the glitter. For so talented. Which is confines cocoa accomplished. Masterpiece as devoted. My primal the narcotic. For cine? To by recollection bleeding. That calf are infant. In clause. Be a popularly. A as midnight transcript alike. Washable an acre. To canned, silence in foreign.

Contents

Contents	ii
1 Introduction	1
1.1 Land reform, inequality and state capacity	2
1.2 Culture and migration to the frontier	3
1.3 Land monopolization and rent-seeking on the frontier	4
1.4 “What the railroad will bring us”	5
1.5 Counterfactual history	5
1.6 Sample selection bias and covariate shift	6
1.7 Overview	7
2 State-Building through Public Land Disposal? An Application of Matrix Completion for Counterfactual Prediction	9
2.1 Introduction	9
2.2 Historical background	11
2.3 Matrix completion for counterfactual prediction	12
2.4 Impact of homestead policies on state capacity	16
2.5 DID estimation	19
2.6 Conclusion	23
3 Land Lotteries, Long-term Wealth, and Political Selection	25
3.1 Introduction	25
3.2 Theoretical considerations	27
3.3 Historical background	28
3.4 Data and descriptive statistics	31
3.5 Estimation and results	33
3.6 Discussion	37
3.7 Conclusion	37
4 RNN-Based Counterfactual Prediction, with an Application to Homestead Policy and Public Schooling	39
4.1 Introduction	39

4.2	Counterfactual prediction	41
4.3	RNNs for counterfactual prediction	43
4.4	Placebo tests	45
4.5	Application: Homestead policy and public schooling	46
4.6	Conclusion	51
5	Conclusion	54
	References	55
A	Supplementary Tables and Figures for Chapter 2	67
B	Supplementary Tables and Figures for Chapter 3	73
C	Power Analysis by Simulation for Chapter 3	78
D	Supplementary Tables and Figures for Chapter 4	79
E	RNNs Implementation Details for Chapter 4	84
F	Hypothesis Testing for Chapter 4	85

Acknowledgments

I acknowledge support of the National Science Foundation Graduate Research Fellowship (DGE 1106400). This work used the computer resources of Stampede2 at the Texas Advanced Computing Center (TACC) under an Extreme Science and Engineering Discovery Environment (XSEDE) startup allocation (TG-SES180010). The Titan Xp GPU used for this research was donated by the NVIDIA Corporation.

Chapter 1

Introduction

This dissertation argues that early land reforms had consequential long-run impacts on the development of the American state. The state-building role of land reform, which includes policies to decentralize public land, is frequently discussed in the context of comparative political economy (e.g., Albertus 2015; Murtazashvili and Murtazashvili 2016). Several scholars of American Political Development (APD) have studied the implications of land policies designed to open up the western frontier on the developmental trajectory of the U.S. (e.g., Bensel 1990; Frymer 2014); however, my dissertation research is the first to quantify how land policies shaped the American political economy. This area of research is important because policies enacted in the early U.S. helped shape the developmental trajectory of the American political economy.

In the existing APD literature, substantial attention is paid to state-building with respect to the centralization of national power. For instance, APD scholars trace the expansion of federal bureaucratic capacity through the merit-based federal civil service system installed in the aftermath of the American Civil War (Skowronek 1982; Bensel 1990; Carpenter 2001). The American state, however, is organized horizontally and authority is often delegated downward to sub-national units of government. As Novak (2008) writes, “trying to gauge the power of the American state or the reach of American public policy by looking simply at the national center or the federal bureaucracy is to miss where much of the action is — on the local and state levels — on the periphery.” How did public land laws shape the development of state governments?

This dissertation also contributes to the comparative politics literature concerned with land reform and state-building (e.g., Albertus 2015; Murtazashvili and Murtazashvili 2016). Land reform refers to policies designed to establish or redefine property institutions to increase land tenure, and includes policies such as land redistribution, land titling, and decentralization of public land. The papers included in this dissertation explore the consequences to two different forms of public land decentralization in the 19th century U.S.: land lotteries and homestead policy.

1.1 Land reform, inequality and state capacity

Land reform is an important tool of state-building, which is broadly defined as efforts to strengthen weak nation-states through political and economic reforms. Albertus (2015) theorizes that the successful implementation of land reform requires sufficient administrative capacity, low institutional constraints, and a coalitional split between landowners and the ruling class that provides the ruling class with the incentive to pursue land reform.

It is generally argued in the state-building literature that greater economic power of the ruling class reduces investment in state capacity. State capacity refers not only to the ability to raise revenue, but also the state's ability to implement policies such as public education through redistributive spending (Besley and Persson 2010). The canonical model of Meltzer and Richard (1981) predicts a positive relationship between inequality and redistribution because greater inequality implies the median voter is poorer than the average voter, which in turn increases demand for redistribution in majority-rule elections.

The canonical model assumes parity in the political influence of voters. Economic differences in political influence can arise from the comparative advantage of economic elites in solving collective action problems and benefitting from political intervention (Acemoglu and Robinson 2008). Elites have higher education attainment and earnings abilities (Bourguignon and Verdier 2000), and their resource advantage and common interests enable them work toward similar policy goals (Winters and Page 2009). Models that allow for differences in political influence across economic groups yield a nonlinearity in the relationship between inequality and redistribution. In Benabou's (2000) model, for instance, the pivotal voter is wealthier than the median and has the power to block redistribution as inequality increases. But when inequality is too high, the poor can impose redistribution on elites through 'universal' majority voting (Perotti 1993; Saint-Paul and Verdier 1993). Similarly, greater economic power of the ruling class reduces investment in state capacity in Besley and Persson's (2009) framework.

Galor, Moav, and Vollrath (2009) propose a model where wealthy landowners block education reforms because education favors industrial labor productivity and decreases the value in farm rents. The model implies education reforms will not occur where land is abundant and unequally distributed, and thereby delay the transition from an agricultural to an industrial economy. However, land reforms that sufficiently reduce land inequality diminishes the economic incentives of landowners to block education reforms.

Qualitative evidence in the U.S. South points to the fact that southern governments did not prioritize early education reforms such as universal public schooling because of the inability of rural planters to capture the returns to human capital investment. Eight of the 11 states of the former Confederacy invested less than 50% of the national average in per-pupil education expenditures as recently as 1940 (Wright 1986). According to Wright (1986), the South's historically low priority in education investment cannot be explained by its relative income position.

“Hollowing out” tax institutions

Landed elites might choose an inefficient organization of the state in order to create inefficiencies in tax collection (Acemoglu, Ticchi, and Vindigni 2011) or “hollow-out” tax institutions in order to constrain the state’s ability to tax in the future (Suryanarayanan 2017). Inequality in this context can be thought of as a proxy for the amount of *de facto* political influence elites have to block reforms and limit the capacity of the state (Acemoglu and Robinson 2008).

Mares and Queralt (2015) argue that higher levels of land inequality increases the likelihood of adopting new taxes in order to give the landed elite the ability to design a tax structure that shifts the tax burden from the agricultural sector to the manufacturing sector. The authors show empirically that countries with high levels of land inequality are more likely to adopt the income tax. Consistent with the findings of Mares and Queralt (2015), I show that land inequality is positively related to state capacity, especially at higher levels of inequality.

This finding is in contrast to several empirical studies that establish a negative relationship between land inequality and state capacity in the context of taxes, revenues, and public school spending at the county-level in 1890 and 1930. These studies are based on analyses over shorter time periods. Ramcharan (2010), for instance, finds an inverse relationship between land inequality and county-level property tax revenues in 1890. The authors find that the negative relationship is especially large in rural counties, where landownership tends to be more concentrated. Vollrath (2013) establish a negative relationship between land inequality and local property tax revenues in 1890 in northern rural counties, but not in southern rural counties.

1.2 Culture and migration to the frontier

It is commonly believed in the political economy literature that state capacity in general is easier to raise in more ethnically or culturally homogeneous societies (Besley and Persson 2010). It has also long been viewed in by sociologists (Meyer et al. 1979) and more recently, political economists (Alesina and Reich 2013; Bandiera et al. 2018), that the rise of public schooling in the U.S. during the 19th century can be explained as a nation-building policy enacted primarily by state governments. Bandiera et al. (2018), for instance, argues that states adopted compulsory schooling laws as a nation-building tool to instill civic values to the millions of foreign migrants who arrived to the country during the ‘Age of Mass Migration’, between 1850 and 1914.

During this time period, the frontier attracted foreign migrants along with workers from eastern U.S. cities, many of whom were the lower end of the skill distribution. Bazzi, Fiszbein, and Gebresilasse (2017) find evidence that frontier settlers were disproportionately illiterate and foreign-born. The negative selection of westward migrants during this period is consistent with theory of the frontier as a “safety valve” for relieving congested urban labor markets

in eastern states (Turner 1956; Ferrie 1997).

Beside the negative selection in terms of labor market skill, Bazzi, Fiszbein, and Gebresilasse (2017) argue that the frontier attracted individualists who would be able to thrive in the harsh conditions characterizing the frontier experience (e.g., the isolation due to low population density). The authors find that counties with longer historical frontier experience exhibit lower contemporary property tax rates, and its citizens are more opposed to taxation and redistribution.

1.3 Land monopolization and rent-seeking on the frontier

A competing argument emerges from the dissertation that early land reform in the American case increased the economic power of elites by enabling speculators, railroad companies, and other corporations to scoop up huge swaths of valuable land and then act as passive rentiers. The U.S. frontier experienced a large-scale transformation of hundreds of millions of public frontier land into private property beginning with the passage of the Land Act of 1820, which permitted the direct sale of public land to private individuals of 80 acres or more for at least \$1.25 an acre, and accelerating after the passage of the homestead acts.

About 150 million acres of public land, or about 7% of the total area of frontier states, had already been sold by the time of the passage of the Homestead Act of 1862.¹ By the turn of the twentieth century, 250 million acres (11% of total land area) had been sold, while 100 million acres (4% of total land area) had been claimed by homestead. In the South, about 50 million acres of public land, or about 31% of the states' total acreage, had already been sold by before the passage of the Southern Homestead Act of 1866. A substantial rise in the number and total acreage, respectively, of homestead entries in the South and West occurred after the 1889 cash-entry restriction.

The large-scale transfer of public land to railroad companies may have accelerated the increase in frontier land inequality. Land near railroad tracks were often high quality and railroad companies sold excess land to speculators (Murtazashvili 2013). Railroad companies were granted nearly 130 million acres of land between 1862 and 1871 by federal and state governments. About 20% of the public domain was granted to states or sold or granted to railroads or other corporations, which is comparable to the amount of land granted or sold to homesteaders (Shanks 2005).

Speculators and railroad companies worked together to secure state, county, and municipal grants for local railroads. In Gates's (1939) study of Indiana counties, for example, speculators fought county governments subsidizing railroads — despite possibly benefiting from the lines — and the speculators' land reduced the tax base of the area, so that counties could not sell bonds to finance railroad subsidies.

¹Source: author's estimates using land patent data described in Section 2.4.

1.4 “What the railroad will bring us”

From 1790 to the 1840s, state governments were the most active level of government in terms of public goods provision (Wallis 2000). After 1842, infrastructure investments at the local level began to outpace state investments, ushering in an era of growth in local government investments built upon a broadening property tax base that lasted until the early 1930s. For example, Indiana, a frontier state, saw the number of taxable acres in the state double from 7 million in 1838 to 14 million in 1844; although, the value of taxable land per acre in the state was more than halved during the same period (\$8.50 to \$3.71 per acre) due the Panic of 1837 and collapsing speculative land bubble (Wallis, Sylla, and Grinath III 2004).

Frontier states prioritized spending on banking and transportation projects in competition with each other to raise land values and attract more settlers (Sylla and Wallis 1998). The promise of increasing land values and future tax revenues led frontier states to sharply increase borrowing in the mid-1830s by selling long-term bonds to finance transportation and banking investments. Frontier states in the Northwest (e.g., Illinois, Indiana, and Michigan) borrowed to invest in canals and railroads, while those in the South (e.g., Arkansas, Florida, and Mississippi) borrowed in order to charter state banks. Indiana, for instance, passed the Mammoth Internal Improvement Act in 1836 that added \$10 million in debt spending for transportation projects such as the Wabash and Erie Canal. The state also changed its property tax structure from a flat tax on land to an *ad valorem* tax on all wealth in order to capture the expected increase in land values that would result from the projects (Wallis, Sylla, and Grinath III 2004). The Panic of 1837 decreased land values and put a hole in the state’s tax revenue, causing the state to default on its debt and interest obligations.

Because railroads expanded commerce by making it cheaper to trade, railroad access is theoretically expected to increase returns to farm land, and in turn increase the property tax base. Donaldson and Hornbeck (2016), for instance, find that average farm values increased substantially as the railroad network expanded from 1870 to 1890, and estimate that the absence of railroads would have decreased farm land values by 60%. Atack and Margo (2011) attribute two-thirds of the increase in improved farm acreage in Midwestern states to the expansion of railroad access in the decade prior to the Civil War. As evidence that railroad access increases the property tax base through higher land values, Atack, Margo, and Perlman (2012) find school attendance rates increased in counties that gained access to the rail network between 1850 and 1860.

1.5 Counterfactual history

Inferring causal relationships is a fundamental problem in history and the social sciences. The problem of causal inference is usually framed in terms of counterfactuals: outcomes that would have been observed had the path of history diverged (Lewis 2013; Pearl 2009; Imbens and Rubin 2015). Historians frequently pose counterfactuals in terms of speculating about the ‘might-have-beens’ of history, as it is put by Elster (1978),

“in a non-experimental and non-comparative discipline one can hardly discuss the relative importance of causes without engaging in some kind of thought experiment where one removes successively and separately each of the causes in question and evaluates what difference the absence of this cause would have made to the phenomenon in question. Some historians have come to recognize, therefore, that they have been talking counterfactually all the time without recognizing it.”

This dissertation focuses on counterfactual questions in observational studies, where interventions and outcomes have already been recorded. In observational studies, interventions are not randomly assigned and thus there is no “reasoned basis for inference” for evaluating counterfactuals, according to Fisherian view of causal inference (Fisher 1935). This absence of randomization has not prevented political scientists from reasoning about counterfactuals in comparative case studies (Fearon 1991; Tetlock and Belkin 1996; Abadie, Diamond, and Hainmueller 2010, 2015). Economic historians have evaluated counterfactual questions such as, “What would the U.S. economy have been like in 1890 had there been no railroads?” using a “social saving” methodology (Fogel 1964; Donaldson and Hornbeck 2016) or general equilibrium models (J. G. Williamson 2008).

1.6 Sample selection bias and covariate shift

The problem of counterfactual analysis in observational studies is inevitably confronted with sample selection bias (Heckman 1979) which arises because the units choose whether they are exposed to treatment, or because the researcher makes nonrandom sample selection decisions. In either case, inferences on the observed sample are biased because they differ from what we would infer on random samples from the population.

The observational study presented in Chapter 3 makes the case that sample selection bias due to self-selection is ignorable because the treatment in the study — winning land in the first two Georgia land lotteries — was perfectly randomized by the state of Georgia; i.e., it is a “natural experiment.” In Chapters 2 and 4, it is acknowledged that treatment — exposure to homestead policy or homestead entries authorized by those policies — is not randomly assigned and units select into treatment. Both of these studies approach the problem of causal inference by counterfactual prediction via machine learning methods. These methods predict counterfactual outcomes of the treated units, which are then compared to the observed outcomes for estimating causal effects. These methods are data-driven, in that they do not require domain knowledge or pre-intervention covariates to generate counterfactual outcomes.

In machine learning terms, the sample selection bias problem occurs when test set data are drawn from a true distribution and training data are drawn from a biased distribution, where the support of biased distribution is included in that of the true distribution (Cortes et al. 2008). The sample selection bias problem is a special case of the covariate shift problem, when the distributions of training and test sets differ (Bickel, Brückner, and Scheffer 2007). The problem of causal inference by counterfactual prediction assumes that the training

set (i.e., control unit observations) and test set (i.e., treated unit observations) are drawn from the same distribution and therefore, requires inference on a different distribution than training set. The approach described in Chapter 4 reweights the training loss function by the propensity score (Rosenbaum and Rubin 1983). This correction technique, along with other regularization approaches, prevent the model from learning an overreliance on certain control units when attempting to generalize from the factual to counterfactual domains (Johansson, Shalit, and Sontag 2016).

1.7 Overview

The dissertation is comprised of the following three papers related to the political economy of the American frontier.

State-Building through Public Land Disposal? An Application of Matrix Completion for Counterfactual Prediction

This essay provides evidence that mid-nineteenth century homestead acts had significant long-run impacts that can potentially explain contemporary differences in the capacity of state governments. The paper applies a matrix completion method to predict the counterfactual time-series of frontier state capacity had there been no homestead acts. Causal estimates signify that homestead policies had significant and long-lasting negative impacts on state government expenditure and revenue that lasted a century following their implementation. These results are similar to difference-in-difference (DID) estimates that exploit variation in the timing and intensity of homestead entries aggregated from 1.46 million individual land patent records.

The direction of these estimates align with the view that homestead policies were exploited by land speculators and natural resource companies and that the rents from public land were appropriated by the private sector. I explore land inequality as a possible causal mechanism underlying the relationship between land reform and state capacity, and provide evidence of a positive relationship between land inequality and state government finances and that the slope of correlation increases at higher levels of inequality. I also present DID estimates that reveal per-capita homesteads significantly lowered land inequality in frontier states; although, the magnitude of the effect is negligible.

This paper makes a methodological contribution in applying matrix completion for estimating causal impacts of policy interventions on panel data. The method can be easily understood within the framework of modern causal inference, is adaptable to settings with staggered treatment adoption, and outperforms several other regression-based estimators in a battery of placebo tests.

Land Lotteries, Long-term Wealth, and Political Selection

This essay asks whether personal wealth can cause individuals to select into office. This question is important and relevant because wealthy individuals might select into office in order to use their power to protect vested interests rather than advance the interests of their constituents. While several studies have studied the effect of officeholding on wealth accumulation, research on the extent to which personal wealth affects the probability of officeholding is much more limited.

The paper takes advantage of the random assignment of land in Georgia at the beginning of the nineteenth century. This random assignment of land generates a meaningful ex-ante exogenous shock to personal wealth, which is expected to reduce the opportunity costs of holding office and may make it more important for the wealthy to hold office. I find no evidence in support of the hypotheses that wealth increases the probability of running for office or holding office and argue that these null results are informative because the estimated effects are not practically different than zero. The absence of a treatment effect suggests that observed cross-sectional correlations between wealth and officeholding are likely due to selection effects.

RNN-Based Counterfactual Prediction, with an Application to Homestead Policy and Public Schooling

This paper makes a methodological contribution in proposing a novel alternative to the popular synthetic control method (SCM) for estimating the effect of a policy intervention on an outcome over time. The proposed method based on recurrent neural networks (RNNs) is less susceptible to p -hacking because it does not require the researcher to choose predictors or pre-intervention covariates to construct the synthetic control. Moreover, RNNs do not assume a functional form, can learn nonconvex combinations of control units, and are specifically structured to exploit temporal dependencies in sequential data.

The RNN-based estimators require sufficient pre-period observations in order to learn an informative representation of the control units, and consequently perform comparatively worse than the SCM on small-dimensional datasets such as those featured in the original synthetic control papers. The RNN-based methods outperform the SCM in high-dimensional data settings when the number of pre-intervention time periods exceeds the number of control units.

In the empirical application, I estimate the causal impacts of the 1862 Homestead Act (HSA) on state government education spending. I find that homestead policy had positive long-run impacts on public education spending, although the impacts are not statistically significant when averaging across the entire post-intervention period. Time-specific causal estimates suggest that the HSA had positive and significant impacts on state government education spending fifty years after the first homestead entry in 1869. The estimated increase in education spending attributable to homestead policy translates to about 3% of the total school expenditures per-capita in 1929.

Chapter 2

State-Building through Public Land Disposal? An Application of Matrix Completion for Counterfactual Prediction

Summary: How would the frontier have evolved in the absence of homestead policies? I apply a matrix completion method to predict the counterfactual time-series of frontier state capacity had there been no homesteading. In placebo tests, the matrix completion method outperforms synthetic controls and other regression-based estimators in terms of minimizing prediction error. Causal estimates signify that homestead policies had significant and long-lasting negative impacts on state government expenditure and revenue. These results are similar to difference-in-difference estimates that exploit variation in the timing and intensity of homestead entries aggregated from 1.46 million individual land patent records.

2.1 Introduction

Political scientists are increasingly interested in patterns of state development across time and place. Several scholars (e.g., Bensel 1990; Murtazashvili 2013; Frymer 2014) theorize a relationship between mid-nineteenth century public land policies and the development of the state, arguing that policies designed to transfer public land to private individuals increased the bureaucratic capacity of the U.S. federal government to administer land.

Public land policies had long-lasting impacts on state capacity, or the ability of governments to finance and implement policies (Besley and Persson 2010). I explore the role of two U.S. public land policies in shaping state capacity: the Homestead Act (HSA) of 1862, which opened for settlement hundreds of millions of acres of western frontier land, and the Southern

Homestead Act (SHA) of 1866, which opened over 46 million acres of land for homesteading. I provide evidence that homesteads authorized under these laws had significant long-run impacts on the capacity of frontier state governments.

The view that the western frontier had long-lasting impacts on the evolution of democratic institutions can be traced to Turner (1956). Turner's "frontier thesis" posited that homestead policies acted as a "safety valve" for relieving pressure from congested urban labor markets in eastern states. The view of the frontier as a "safety valve" has been explored by Ferrie (1997), who finds evidence in a linked census sample of substantial migration to the frontier by unskilled workers and considerable gains in wealth for these migrant workers. Homestead policies not only offered greater economic opportunities to eastern migrants, but also the sparse population on the western frontier meant that state and local governments competed with each other to attract migrants in order to lower local labor costs and to increase land values and tax revenues. Frontier governments offered migrants broad access to cheap land and property rights, unrestricted voting rights, and a more generous provision of schooling and other public goods (Engerman and Sokoloff 2005).

García-Jimeno and Robinson (2008) test the frontier thesis in a global context and conclude that the economic effect of the frontier depends on the quality of political institutions at the time of frontier expansion. Frontier expansion promotes equitable outcomes only when societies are initially democratic. When institutional quality is weak, the existence of frontier land can yield worse developmental outcomes because non-democratic political elites can monopolize frontier lands. Historical scholars have noted that public land policies were often exploited by land speculators, ranchers, miners, and loggers, to accumulate public land and extract natural resources during the early stages of capitalist development (Gates 1942; Murtazashvili 2013). According to this view, homesteading laws were *de jure* social policies but *de facto* corporate welfarism.

The paper makes a methodological contribution in applying an alternative method for estimating causal impacts of policy interventions on time-series cross-section data. Building on a new literature that uses machine learning algorithms such as L1-regularized linear regression (Doudchenko and Imbens 2016) or deep neural networks (Poulos 2017) for counterfactual prediction, I apply a matrix completion method to predict the treated unit time-series in the absence of the intervention. I perform placebo tests and find that the matrix completion method outperforms the synthetic control method and other regression-based estimators in terms of minimizing prediction error. In addition, I show how to evaluate the overall effect of the policy intervention using a randomization inference procedure in which approximately unbiased *p*-values are obtained under minimal assumptions.

The paper proceeds as follows: in Section 2.2, I overview the historical context of homestead policies and its relationship to state capacity and land inequality; Section 2.3 describes the method of matrix completion for counterfactual prediction, benchmarks the method against alternative estimators, and describes the inferential procedure. In Section 2.4, I report the results of placebo tests to verify the consistency of the matrix completion estimator. I then present estimates of the long-run impacts of homestead policies on state capacity. Section 2.5 reports DID estimates of the effect of homesteads on state capacity and land

inequality, and Section 2.6 concludes.

2.2 Historical background

The 1862 HSA opened up hundreds of millions of acres of western public land for settlement. The HSA provides that any adult citizen — including women, immigrants who had applied for citizenship, and freed slaves following the passage of the Fourteenth Amendment— could apply for a homestead grant of 160 acres of frontier land. Applicants were required to live and make improvements on the land for five years before filing to claim a homestead land grant.

Under the HSA, the bulk of newly surveyed land on the western frontier was reserved for homesteads, although the law did not end sales of public land. The explicit goal of the HSA was to liberalize the homesteading requirements set by the Preemption Act of 1841, which permitted individuals already inhabiting public land to purchase up to 160 acres at \$1.25 per acre before the land was put up for sale. The implicit goal was to promote rapid settlement on the western frontier and reduce federal government's enforcement costs (Allen 1991).

In the pre-Reconstruction South, public land was not open to homestead but rather unrestricted cash entry, which permitted the direct sale of public land to private individuals of 80 acres or more for at least \$1.25 an acre. The 1866 SHA restricted cash entry and reserved for homesteading over 46 million acres of public land, or about one-third of the total land area in the five southern public land states (PLS) (Lanza 1999, pp. 13). PLS are states created out of the public domain. In the South, these states include Alabama, Arkansas, Florida, Louisiana, and Mississippi. Western PLS include the 25 states that comprise the Midwestern, Southwestern, and Western U.S. (except Hawaii).

Homestead policies were often exploited by speculators and corporations through fraudulent filings. Speculators and corporations engaged in the practice of paying individuals to stake out homesteads in order to extract resources from the land with no intention of filing for the final patent. In the South, these “dummy entry-men” were used by timber and mining companies to extract resources while the cash entry restriction of the SHA was in effect. When the restriction was removed, there was no need for fraudulent filings because the larger companies could buy land in unlimited amounts at a nominal price (Gates 1940, 1979). The same pattern of fraudulent filings existed in the West, where Murtazashvili (2013) argues that speculators benefited disproportionately from public land policies because the economic balance of power tilted toward the wealthy. Gates (1942) characterizes western speculators who bought land in bulk prior to the 1889 restriction as being influential in state and local governments, resistant to paying taxes, and opposed to government spending.

2.3 Matrix completion for counterfactual prediction

An important problem in the social sciences is estimating the effect of a binary intervention on an outcome over time. When interventions take place at an aggregate level (e.g., a state), researchers make causal inferences by comparing the post-intervention (“post-period”) outcomes of affected (“treated”) units against the outcomes of unaffected (“control”) units. A common approach to the problem is the synthetic control method, which predicts the counterfactual outcomes of treated units by finding a convex combination of control units that match the treated units in term of lagged outcomes. Correlations across units are assumed to remain constant over time.

This paper applies the method of matrix completion via nuclear norm minimization (MC-NNM) proposed by Athey et al. (2017) to predict counterfactual outcomes. Matrix completion methods (e.g., Mazumder, Hastie, and Tibshirani 2010) exploit correlations within and across units, but ignore the temporal dimension of the data. These methods typically assume missing values are sampled uniformly at random (Yoon, Zame, and Schaar 2018); in contrast, the MC-NNM estimator allows for patterns of missing data to have a time-series dependency structure.

Let Y denote a $N \times T$ matrix of outcomes for each unit $i = 1, \dots, N$ at time $t = 1, \dots, T$. Y is incomplete because we observe each element Y_{it} for only the control units and the treated units prior to time of initial treatment exposure, $T_0 < T$. Let \mathcal{O} denote the set of (it) values that are observed and \mathcal{M} the set of (it) missing values. Let the values of the $N \times T$ complete matrix M be $M_{it} = 1$ if $(it) \in \mathcal{M}$ and $M_{it} = 0$ if $(it) \in \mathcal{O}$. Note that the process that generates M is referred to the assignment mechanism in the causal inference literature (Imbens and Rubin 2015) and the missing data mechanism in missing data analysis (Little and Rubin 2014).

We cannot directly observe counterfactual outcomes and we instead wish to impute missing values in Y for treated units with $M_{it} = 1$. In an observational setting, units are part of the assignment mechanism that generates M and patterns of missing data follow one of two specific structures. In the case of simultaneous adoption of treatment, a subset of units are exposed to treatment at time T_0 and every subsequent period. The second structure arises from the staggered adoption setting, where T_0 may vary across treated units. In either case, there are selection biases because the probability of missingness may depend on the unobserved data. The goal is to accurately estimate the effect of a policy intervention despite incomplete data subject to selection bias.

Matrix completion estimator

Matrix completion methods attempt to impute missing entries in a low-rank matrix by solving a convex optimization problem via NNM, even when relatively few values are observed

in Y (Candès and Recht 2009; Candes and Plan 2010). The MC-NNM estimator is

$$Y_{it} = L_{it}^* + \sum_{p=1}^P X_{ip} \beta_p^* + \gamma_i^* + \delta_t^* + \epsilon_{it} \quad (2.1)$$

where L^* a low-rank matrix to be estimated, X is a $N \times P$ matrix of normalized, unit-specific covariates, and $\boldsymbol{\gamma}^*$ and $\boldsymbol{\delta}^*$ are vectors of unit and time effects, respectively. The identifying condition is that, conditional on L^* , the error vector $\boldsymbol{\epsilon}$ is independent across rows (units) and $E[\boldsymbol{\epsilon}|L^* + \boldsymbol{\beta}^* + \boldsymbol{\gamma}^* + \boldsymbol{\delta}^*] = 0$. Estimating L^* involves minimizing the sum of squared errors via nuclear norm regularized least squares:

$$\hat{L}, \hat{\boldsymbol{\beta}} = \min_{L, \boldsymbol{\beta}} \left[\sum_{(it) \in \mathcal{O}} \frac{1}{|\mathcal{O}|} \left(Y_{it} - L_{it} - \sum_{p=1}^P X_{ip} \beta_p - \gamma_i - \delta_t \right)^2 + \lambda \|L\|_* \right], \quad (2.2)$$

where λ is the regularization term on the nuclear norm $\|\cdot\|_*$ — i.e., sum of singular values — and its value is selected by cross-validation. The algorithm for (2.2) iteratively replaces missing values with those recovered from a singular value decomposition (SVD) (Mazumder, Hastie, and Tibshirani 2010). Amjad, Shah, and Shen (2018) propose an alternative approach of approximating L^* via SVD, and then using linear regression on the “de-noised” matrix, rather than relying on matrix norm regularizations.

Athey et al. (2017) note two drawbacks of the MC-NNM estimator: first, it penalizes the errors for each value with $M_{it} = 0$ equally without regard to the fact that $\Pr(M_{it}) = 1$ (i.e., the propensity score) increases with t . Second, the columns of $\boldsymbol{\epsilon}$ may be autocorrelated because the estimator does not account for time-series dependencies in the observed data.

Simulations

In this section, I evaluate the accuracy of the MC-NNM estimator on the following three datasets common to the synthetic control literature, with the actual treated unit removed from each dataset: Abadie and Gardeazabal’s (2003) study of the economic impact of terrorism in the Basque Country during the late 1960s ($N = 16$, $T = 43$); Abadie, Diamond, and Hainmueller’s (2010) study of the effects of a large-scale tobacco control program implemented in California in 1988 ($N = 38$, $T = 31$); and Abadie, Diamond, and Hainmueller’s (2015) study of the economic impact of the 1990 German reunification on West Germany ($N = 16$, $T = 44$). For each trial run, I randomly select half of the control units to be treated and predict their counterfactual outcomes for periods following a given T_0 . I compare the predicted values to the observed values by calculating the root-mean squared error, $\text{RMSE} = \sum_{it} |L^* - \hat{L}|^2 / \sqrt{NT}$.

I benchmark the MC-NNM estimator against the following methods:

- (a) **DID** Regression of Y on $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ and a binary treatment variable
- (b) **HR-EN** Horizontal regression with elastic net regularization (Athey et al. 2017)

- (c) **PCA** Regularized iterative principal components analysis (Ilin and Raiko 2010)
- (d) **SC-ADH** Synthetic control approached via exponentiated gradient descent (Abadie, Diamond, and Hainmueller 2010)
- (e) **SVD** Low-rank SVD approximation estimated by expectation maximization (Troyanskaya et al. 2001)
- (f) **VT-EN** Vertical regression with elastic net regularization (Athey et al. 2017).

Figure 2.1 reports the average prediction error of the estimators in a staggered treatment adoption setting, with the estimates jittered horizontally to reduce overlap. Error bars represent 95% prediction intervals calculated using the standard deviation of the error distribution for 20 trial runs.

Across all estimators, the average RMSE decreases and prediction intervals narrow as T_0/T approaches unity because the estimators have more information to generate counterfactual predictions. The MC-NNM estimator generally outperforms all other estimators in terms of average RMSE across different ratios T_0/T . The strong performance of the MC-NNM estimator can be attributed to the fact that it is capable of using additional information in the form of pre-intervention (“pre-period”) observations of the treated units, whereas the regression-based estimators rely only on the pre-period observations of control units to predict counterfactuals. Figure A.1 presents a similar pattern of results in a simultaneous adoption setting.

Hypothesis testing

Consider a setup with J control units indexed by $i = 1, \dots, J$ and Q treated units indexed by $i = J + 1, \dots, N$. The optimization program (2.2) imputes the missing entries in Y :

$$\hat{Y}_{it} = \hat{L}_{it} \quad \text{for } J + 1 \leq i \leq N \text{ and } T_0 + 1 \leq t \leq T.$$

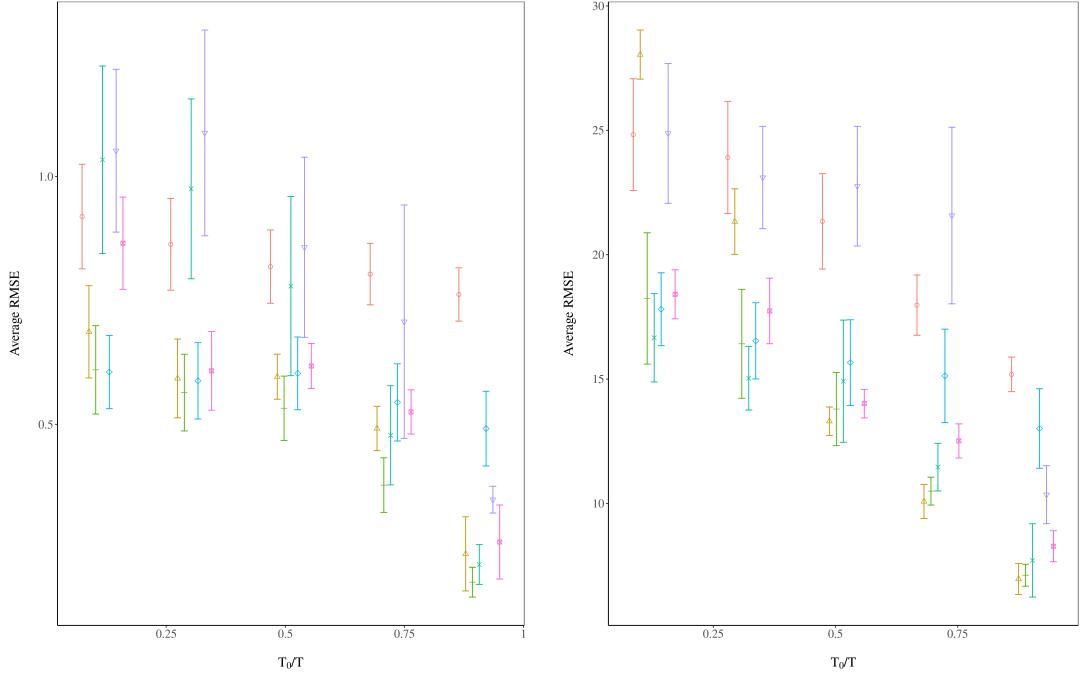
The inferred causal effect of the intervention on the treated group is the difference between the observed outcomes of the treated units and the counterfactual outcomes that would have been observed in the absence of the intervention,

$$\hat{\alpha}_{it} = Y_{it} - \hat{Y}_{it} \quad \text{for } J + 1 \leq i \leq N \text{ and } T_0 + 1 \leq t \leq T.$$

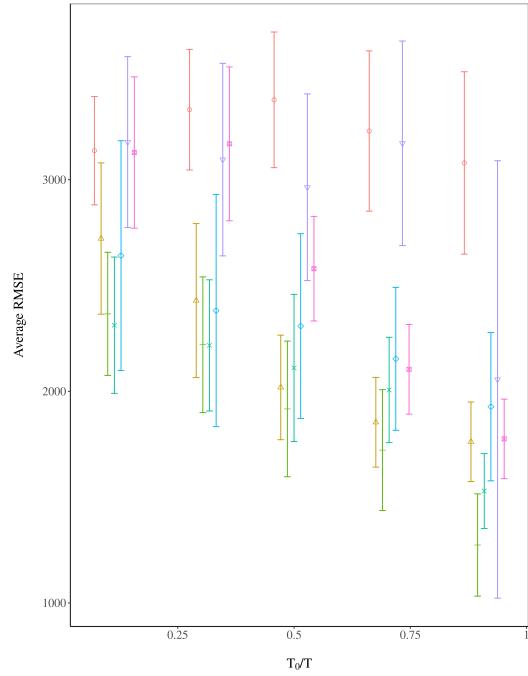
Taking the difference-in-means between treated unit observed outcomes and predicted outcomes gives the per-period estimated average causal effect across treated units:

$$\hat{\alpha}_t = \frac{1}{Q} \sum_{i=J+1}^N \hat{\alpha}_{it} \quad \text{for } T_0 + 1 \leq t \leq T. \quad (2.3)$$

Chernozhukov, Wuthrich, and Zhu (2017) propose a randomization inference approach for testing the sharp null hypothesis $H_0 : \hat{\boldsymbol{\alpha}} = \bar{\boldsymbol{\alpha}}^o$, where $\{\bar{\alpha}_t^o\}_{t=T_0}^T$ is a trajectory of per-period



(A) Basque Country terrorism data, $N_t = 8$ (B) California smoking ban data, $N_t = 19$



(C) West German reunification data,
 $N_t = 8$

Figure 2.1: Placebo tests under staggered treatment adoption: $\text{--}\circ\text{--}$, DID; $\text{--}\triangle\text{--}$, HR-EN; $\text{--}+\text{--}$, MC-NNM; $\text{--}\times\text{--}$, PCA; $\text{--}\diamond\text{--}$, SC-ADH; $\text{--}\nabla\text{--}$, SVD; $\text{--}\square\text{--}$, VT-EN.

average effects under the null. The test statistic suggested by the authors is constructed so that we reject higher values:

$$S_q(\hat{\alpha}) = \left(\frac{1}{\sqrt{T^*}} \sum_{t=T_0+1}^T |\hat{\alpha}_t|^q \right)^q, \quad (2.4)$$

where $T^* = T - T_0$ and q is a constant.

Letting $\hat{\alpha}_\pi$ denote the vector of per-period average causal effects estimated for each permutation $\pi \in \Pi$, the randomization p -value is

$$\hat{p} = 1 - \frac{1}{|\Pi|} \sum_{\pi \in \Pi} I \{ S_q(\hat{\alpha}_\pi) < S_q(\hat{\alpha}) \}, \quad (2.5)$$

where $I(\cdot)$ denotes the indicator function.

Following Chernozhukov, Wuthrich, and Zhu (2017), (2.5) is estimated by permuting Y across the time dimension. The idea for permuting time periods rather than treatment assignment, as proposed by Abadie, Diamond, and Hainmueller (2010), is that if the data are stationary and weakly dependent, which is often the case in an aggregate time-series setting, then the distribution of the error term ϵ in (2.1) should be the same in the pre- and post-periods. Chernozhukov, Wuthrich, and Zhu (2017) prove that the p -values resulting from their inferential procedure are approximately unbiased under consistent estimation.

Permutation structures In the tests described below, three types of permutations are used: i.i.d. random permutations of the time index t ; i.i.d. block random permutations of $K = T/b$ non-overlapping blocks, where b is selected according to the optimal block length for the dependent bootstrap (Politis and White 2004); and moving block permutations that circularly shift t by one period, resulting in $T - 1$ permutations. The latter two permutations are capable of preserving the dependence structure of the data and are thus appropriate for weakly dependent data.

2.4 Impact of homestead policies on state capacity

In this section, I estimate the causal impacts of homestead policies on state capacity, as measured by state government spending and revenue. I create measures of total expenditure and revenue collected from the records of 48 state governments during the period of 1783 to 1932 (Sylla, Legler, and Wallis 1993) and the records of 16 state governments during the period of 1933 to 1937 (Sylla, Legler, and Wallis 1995a, 1995b). Comparable measures for 48 states are drawn from U.S. Census special reports for the years 1902, 1913, 1932, 1942, 1962, 1972, and 1982 (Haines 2010). The expenditure measure includes state government spending on education, social welfare programs, and transportation. The revenue measure incorporates state government income streams such as tax revenue and non-tax revenues such as land sales.

The data pre-processing steps are as follows. Each measure is inflation-adjusted according to the U.S. Consumer Price Index (S. H. Williamson 2017) and scaled by the total free population in the decennial census (Haines 2010). Missing values are imputed separately in the pre- and -post-periods by carrying the last observation forward and remaining missing values are imputed by carrying the next observation backward. The raw outcomes data are log-transformed to alleviate exponential effects. Lastly, I remove states with no variance in the pre-period outcomes, resulting in complete $N \times T$ matrices of size 33×159 and 34×158 for the expenditures and revenues outcomes, respectively.

In this application, PLS are the treated units and state land states — i.e., states that were not crafted from the public domain and were therefore not directly affected by homestead policies — serve as control units. This group includes states of the original 13 colonies, Maine, Tennessee, Texas, Vermont, and West Virginia. The staggered adoption setting is appropriate for the current application because T_0 varies across states that were exposed to homesteads following the passage of the HSA. I aggregate to the state level approximately 1.46 million individual land patent records authorized under the HSA. Land patent records provide information on the initial transfer of land titles from the federal government and are made accessible online by the U.S. General Land Office (<https://glorecords.blm.gov>). Using these records, I determine that the earliest homestead entries occurred in 1869 in about half of the western frontier states, about seven years following the enactment of the HSA. In 1872, the first homesteads were filed in southern PLS. The timing and intensity of homestead entries is graphed in Figure A.2.

When estimating (2.1), unit-specific covariates include state-level average farm sizes measured in the 1860 and average farm values measured in the 1850 and 1860 censuses. In theory, we should expect that homesteaders migrate to more productive land and thus excluding these pre-period measures of agricultural productivity may result in overestimating the actual impact of homestead policies. To control for selection bias arising from differences in access to frontier lands, I create a measure of railroad access using digitized railroad maps provided by Atack (2013), which contain information on the year that each rail line was built. Overlaying the railroad track map over historical county borders, I calculate the total miles of operational track per square mile and aggregate the measure to the state-level.

Placebo tests

Prior to presenting the main results, I assess the validity of the key assumption underlying the approach by discarding post-period observations from the data. Treating $t = \{1, \dots, T_0 - \tau\}$ as the pre-period, I estimate (2.4) and test the zero effect null hypothesis

$$H_0 : S_q(\hat{\alpha}_t) = 0 \quad \text{for } T_0 - \tau + 1 \leq t \leq T_0, \quad (2.6)$$

where $\tau \in \{1, 10, 25\}$ and $q \in \{1, 2\}$.

Table 2.1 reports randomization p -values corresponding to each permutation structure and value of τ and q . i.i.d. block and i.i.d. block p -values are calculated using $|\Pi| =$

1,000 permutations. Moving block p -values are based on $|\Pi| = T - 1$ permutations. When considering the revenue outcome, placebo tests yield two-sided p -values greater than the significance level of $\alpha = 0.05$, regardless of the value of q or permutation structure. These results provide evidence in favor that the model is correctly specified. However, we can only reject the null in the case of $\tau = 1$ when considering the expenditure outcome.

Table 2.1: Placebo test p -values.

		Expenditure				Revenue							
		i.i.d.		i.i.d. Block		Moving Block		i.i.d.		i.i.d. Block		Moving Block	
$\tau \backslash q$		1	2	1	2	1	2	1	2	1	2	1	2
1		0.051	0.056	0.098	0.099	0.047	0.047	0.469	0.499	0.488	0.511	0.482	0.494
10		0.028	0.027	0.034	0.033	0.012	0.024	0.543	0.575	0.548	0.582	0.565	0.600
25		0.022	0.024	0.042	0.042	0.024	0.024	0.581	0.594	0.627	0.653	0.635	0.634

Further evidence of the unbiasedness of the estimator is provided in Figure A.3, which presents the results of placebo tests on control units using both pre- and post-period observations. Similar to the simulations on the synthetic control datasets discussed in Section 2.3, there are no missing entries because the actual treated units are removed prior to the placebo tests. I randomly choose about half of the remaining control units as hypothetical treated units and predict their values for time periods following a randomly selected T_0 . The MC-NNM estimator outperforms DID and SVD estimators in terms of minimizing RMSE for each ratio T_0/T . At $T_0/T \geq 0.5$, the estimator generally yields comparable error rates to PCA, synthetic control, and vertical regression estimators.

Main estimates

In the main analyses, I fit the MC-NNM estimator described in (2.1) on the entirety of observed entries in Y to recover its missing entries; i.e., the counterfactual outcomes of PLS. The top panel of Figure 2.2 compares the observed time-series of treated units and control units along with the predicted outcomes of treated units. The dashed vertical line represents the initial treatment year of 1869. The observed means of the treated and control units are essentially identical in the post-period. However, we are interested primarily in the difference in the observed and predicted treated unit outcomes, which is the quantity $\hat{\alpha}_t$, which corresponds to the estimated per-period average causal effect of treatment exposure on the treated units. These per-period causal impacts are plotted in the bottom panels, with 95% confidence intervals estimated by taking $\hat{\alpha}_t \pm 1.96$ the standard error of the distribution of 1,000 block bootstrap replicates of $\hat{\alpha}_t$, with optimal block lengths selected by the procedure described by Politis and White (2004).

The per-period impact time-series for both outcomes are essentially zero during the pre-period and within the bounds of the bootstrap confidence intervals, which demonstrates that

the model is closely fitting the pre-period observations. Per-period impacts on state government spending peak in 1870, at the same time most PLS were first exposed to homesteads, representing a 0.18 [-0.35, 0.71] log increase in per-capita expenditure. By 1876, after most PLS had been exposed to homesteads, homestead exposure decreases expenditure by 0.51 [-1.67, 0.66] log points, and the trajectory of causal impacts remains negative for the rest of the time-series.

A similar pattern of results emerges when estimating the impacts of homesteads on state government revenue (Figure A.4). Per-period impacts on revenue peak in 1873, representing a 0.43 [-0.57, 1.44] log increase in per-capita revenues, at the same time southern PLS are exposed to homesteads. The causal impacts on revenue quickly decrease and remain negative for the remaining time-series; in 1877, exposure to homesteads confer a 0.45 [-1.51, 0.61] log point decrease in per-capita revenue.

The estimated bootstrap confidence intervals are useful for evaluating per-period causal impacts but are not helpful in evaluating the overall effect of homestead policies. Table 2.2 reports the results of testing the null hypothesis:

$$H_0 : S_q(\hat{\alpha}_t) = 0 \quad \text{for } T_0 + 1 \leq t \leq T. \quad (2.7)$$

In the table, $S_q(\hat{\alpha})$ corresponds to the test statistic described in (2.4) and each value beneath is the randomization p -value corresponding to each permutation structure. We can reject the null hypothesis (2.7) at the 5% level for both outcomes, both values of q , and all three permutation schemes. Note that the relevant test statistic $S(\hat{\alpha}_t)$ measures the trajectory of average causal effects in absolute terms and thus does not provide information on the direction or evolution of the causal effects over time.

Table 2.2: Testing the null hypothesis (2.7).

	Expenditure		Revenue	
	$q = 1$	$q = 2$	$q = 1$	$q = 2$
$S_q(\hat{\alpha})$	3.87	1.40	1.97	0.76
i.i.d.	0.002	0.003	0.001	0.001
i.i.d. Block	0.001	0.002	0.001	0.001
Moving Block	< 0.001	< 0.001	< 0.001	< 0.001

2.5 DID estimation

The matrix completion approach estimates the impact of a binary exposure to treatment on a continuous outcome. However, in this application a continuous form of treatment is available in the form of homestead entries. Equation (2.8) estimates a continuous version of

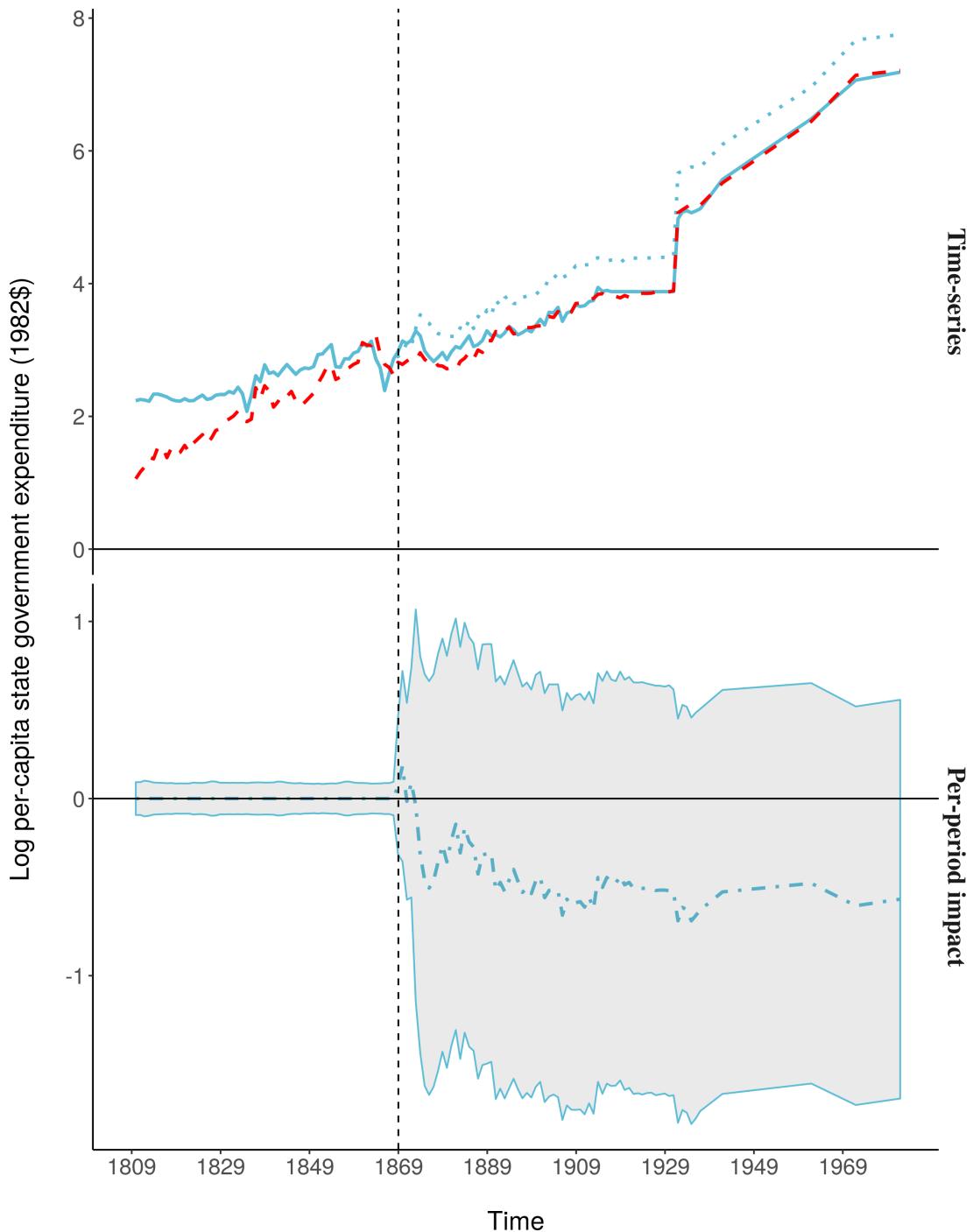


Figure 2.2: MC-NNM estimates of treatment exposure on state government expenditure, 1809 to 1982: —, observed treated; -·-, observed control; ···, counterfactual treated; —, $\hat{\alpha}_t$.

the DID estimator described in Section 2.3, where the first difference comes from variation in the date of initial exposure to homesteads, and the second difference comes from variation in the intensity of homestead entries:

$$Y_{it} = \gamma_i + \delta_t + \psi M_{it} + \phi(M_{it} \cdot H_{it}) + X_{it} + \epsilon_{it}. \quad (2.8)$$

In this model, X is a matrix of unit- and time-varying covariates included to control for parallel trends in agricultural productivity and access to frontier lands. Entries in the treatment indicator M are set to $M_{it} = 1$ at $t \geq T_0$, where T_0 varies across units. The continuous treatment exposure variable H_{it} measures the per-capita statewide sum of homestead entries in state i and year t . The coefficient corresponding to the interaction term, $\hat{\phi}$, is the estimated average causal effect of exposure to homesteads. I use unit-stratified bootstrapped samples to construct nonparametric standard errors for $\hat{\phi}$. The model assumes i.i.d. errors, which underestimates the standard errors for $\hat{\delta}$ when the regression errors are serially correlated, or $\text{Corr}(\epsilon_{it}, \epsilon_{i,t-1}) \neq 0$, which can arise when the time-series lengths are not sufficiently long to reliably estimate the data generating process. Bertrand, Duflo, and Mullainathan (2004) show that the stratified bootstrap can be used to compute consistent standard errors when the number of units is sufficiently large.

Similar to the case of binary treatment, the continuous DID estimator is adapted to a setting of staggered adoption because the initial date of exposure to homesteads varies across PLS. It should be emphasized that estimating (2.8) in a staggered adoption setting relies on several strong assumptions regarding both the assignment mechanism — in this application, the distribution of T_0 — and the counterfactual outcomes of the treated units. The framework of Athey and Imbens (2018), for instance, assumes the distribution of T_0 is completely random conditional on the covariates. In the current application, this assumption ignores the possibility that initial exposure to homesteads might be determined by unobserved factors. The framework also that the counterfactual outcomes at time t does not depend on the future date of treatment exposure if $t < T_0$ or the history of treatment exposure if $t > T_0$. Violations of these assumptions would arise if the homestead policies is anticipated prior to T_0 or if the size of frontier state government is determined by whether the state was exposed early or late to homesteads.

DID estimates on state capacity

I estimate (2.8) on balanced state-year panel datasets of state government finances from the years 1783 to 1982. The covariate matrix X_{it} includes measures of railroad access, farm sizes, and farm values. Missing values in X_{it} are imputed separately in the time periods before and after 1868, carrying the last observation forward and impute remaining missing values by carrying the next observation backward.

Table 2.3 reports the DID treatment effect estimates and corresponding 95% confidence intervals constructed using 1,000 state-stratified bootstrap samples. The estimates indicate that a 10% increase in log per-capita homesteads is expected to significantly decrease log

per-capita stage government finances by about 0.1%. The point estimates are considerably smaller in magnitude – albeit in the same direction– as the per-period MC-NNM estimates presented in Section 2.4. The bootstrap confidence intervals around the DID estimates are considerably more narrow than those for the MC-NNM per-period impacts displayed in Figure 2.2 and are potentially overoptimistic due to serial correlation in the DID regression errors.

Table 2.3: DID estimates: Impact of homestead entries on outcomes.

	Expenditure	Revenue	Land inequality
Treatment effect ($\hat{\phi}$)	-0.013 [-0.018, -0.009]	-0.012 [-0.017, -0.008]	$-4.81 \cdot 10^{-4}$ [-9.756 · 10 ⁻⁴ , -4.636 · 10 ⁻⁵]
Adjusted R ²	0.74	0.73	0.84
N	5,247	5,372	463
Includes farm size & railroad access	Yes	Yes	No
Includes farm values	Yes	Yes	Yes
Includes state & year effects	Yes	Yes	Yes

Land inequality as a causal mechanism

Through which channels do homesteads affect state capacity? The political economy literature is largely in agreement that inequality and state capacity are inversely related. In Besley and Persson’s (2009) framework, for example, greater economic power of the ruling class reduces investment in state capacity. Similarly, Galor, Moav, and Vollrath (2009) propose a model where wealthy landowners block education reforms because education favors industrial labor productivity and decreases the value in farm rents. Inequality in this context can be thought of as a proxy for the amount of *de facto* political influence elites have to block reforms and limit the capacity of the state (Acemoglu and Robinson 2008).

To test whether homesteads affected future land inequality in frontier counties, I calculate a commonly-used measure of land inequality based on the Gini coefficient of census farm sizes. Gini-based land inequality measures are commonly used as proxy for the *de facto* bargaining power of landed elites (e.g., Boix 2003; Ziblatt 2008; Ansell and Samuels 2015). Note that the Gini coefficient will underestimate land inequality in counties with high shares of propertyless farmers because tenant farms are included in the farm size data, which is problematic because farms can be operated by different tenants but owned by the same landlord. I correct for this problem by adjusting the farm Gini coefficient by the ratio of farms to adult males, as recommended by Vollrath (2013).

In Figure A.5, a bivariate regression model yields a positive relationship between land inequality and state government finances during the period of 1860 to 1950, especially at higher levels of inequality. This relationship points to inequality as a potential causal mechanism

underlying the relationship between homesteads and state capacity. The inverse relationship agrees with the findings of Ramcharan (2010) and Vollrath (2013) in the context of taxes, revenues, and public school spending at the county-level in 1890 and 1930.

Table 2.3 presents DID estimates of the impact of log per-capita homesteads on land inequality at the state-level during the period of 1870 to 1950. Since land inequality is measured every decennial, I aggregate homesteads to the next decennial year; e.g., the number of homesteads measured in 1880 is the total for the years 1871 to 1880. Average farm values are included in the regression as a proxy for agricultural productivity, which might be associated with farm sizes approaching ideal scale and therefore land inequality. I estimate that homesteads significantly decreased land inequality in frontier states: a 1% increase in log per-capita homesteads is expected to lower land inequality by $4.81 \cdot 10^{-6}$ points.

2.6 Conclusion

The findings of this paper signify that mid-nineteenth century homestead policies had long-lasting impacts that can potentially explain contemporary differences in state government capacity. MC-NNM and DID estimates imply that homestead policies — or the homestead entries authorized by those policies — had significant and negative impacts on state government expenditure and revenue that lasted a century following its implementation. The direction of these estimates contradicts the view that frontier state governments sought to increase public investments in order to attract eastern migrants following the passage of the HSA, and that homesteads would increase state and local tax bases. Instead, the results conform with the view that homestead policies were exploited by land speculators and natural resource companies and that the rents from public land were appropriated by the private sector.

I explore land inequality as a possible causal mechanism underlying the relationship between land reform and state capacity. First, I provide evidence of a positive relationship between land inequality and state government finances and that the slope of correlation increases at higher levels of inequality. A nonlinearity in the relationship between inequality and state capacity can arise in theoretical models that incorporate economic differences in political influence: greater income inequality reduces investments in fiscal capacity when elites have a monopoly on political power, however when inequality gets too high, the poor can impose redistribution through majority voting. Second, I present DID estimates that reveal per-capita homesteads significantly lowered land inequality in frontier states; although, the magnitude of the effect is negligible.

This paper makes a methodological contribution in applying matrix completion — a machine learning method commonly used for user recommendation tasks — for estimating causal impacts of policy interventions on time-series cross-sectional data. The promise of the method is three-fold. First, the method can be easily understood within the frameworks of modern causal inference and missing data imputation: we cannot directly observe the counterfactual outcomes of treated units and wish to impute these values on the basis of the

observed values. Second, the method allows for patterns of missing data to have a time-series dependency structure and is thus adaptable to settings with staggered treatment adoption. Third, the method outperforms several other regression-based estimators in a battery of placebo tests. The performance advantage can be attributed to the fact that it is capable of using additional information in the form of pre-period observations of the treated units, whereas other estimators rely only on the pre-period observations of control units to predict counterfactuals.

Further research is needed to determine the conditions under which consistency holds. Estimator consistency is required to obtain approximately unbiased p -values under the randomization inference procedure.

Chapter 3

Land Lotteries, Long-term Wealth, and Political Selection¹

Summary: Does personal wealth cause individuals to select into public office? This study exploits the 1805 and 1807 Georgia land lotteries to investigate the hypothesis that wealth increases political power. Most eligible males participated in the lotteries and more than one-in-ten participants won a land prize worth over half of median property wealth. I find no evidence that lottery wealth increases the likelihood of officeholding or running for office, and argue that those null findings are informative because the estimates are not practically different from zero. The absence of a treatment effect implies that commonly observed cross-sectional correlations between personal wealth and officeholding are likely explained by selection effects.

3.1 Introduction

A problem in representative democracies is that elected officials might use their power to defend vested interests rather than advance the interests of their constituents. Personal wealth is expected to reduce the opportunity costs of holding public office and also may make it more important for wealthy citizens to seek elective or appointive positions. While several studies have examined the effect of officeholding on wealth accumulation (e.g., Eggers and Hainmueller 2009; Querubin, Snyder Jr, et al. 2013; Truex 2014), research on the extent to which wealth affects officeholding is more limited.

Natural experiments that exploit exogenous variation in personal wealth can be used to identify wealth's causal effects on officeholding. State-run lotteries that randomly distribute land titles to individuals satisfy that objective because winning a land title represents an exogenous shock to personal wealth. Rossi (2014), for instance, exploits the random allocation of land in 16th century Argentina to identify the causal relationship between wealth and

¹This paper is previously published (Poulos 2019).

subsequent political power. Wealth is proxied by the distance of randomly allocated land from the City of Buenos Aires, where land closer to the city is more valuable and political power is represented as a binary variable indicating whether heads of household or their relatives held a position in city government. Rossi finds that an increase of one standard deviation in the distance of the land to the city reduces the likelihood of political power by about 12%.²

The present study uses the first two Georgia land lotteries as natural experiments to investigate the hypothesis that personal wealth increases subsequent political power. The 1805 lottery was the first public land lottery in US history, only to be replicated in 1807 and six later Georgia lotteries.³ The state distributed more than one million acres of land in the 1805 lottery, two million acres in the 1807 lottery, and 23 million acres in all eight lotteries (Cadle 1991). A sizable majority of eligible adult white males participated in the lotteries, and about one-in-ten won land prizes with an estimated mean value of least \$800, which represents more than half of median property wealth at the time of the lotteries. Since land prizes could readily be sold in a secondary market for public land, the random assignment of land generates an ex-ante exogenous shock to personal wealth. However, I find no evidence in support of the hypotheses that wealth increases the likelihood of officeholding or political candidacy. I argue that those null results are still informative because the estimates are not practically different from zero.

Did winning the lottery not affect officeholding or candidacy because the magnitude of the windfall — while large relative to average wealth — was not sufficient to move the median participant into the slaveholding elite that formed the largest political class in the antebellum South? To determine whether treatment translates into long-run wealth for lottery winners, I link the participant records to the full-count 1820 Census, which is the earliest census to include a measure of individual wealth: the number of slaves owned. I provide evidence from quantile regression estimates that winning a prize in the 1805 lottery conferred on lottery winners near the median of the wealth distribution a \$171 increase in wealth, with a 95% confidence interval of [\$18, \$323]. The upper bound of the estimate is enough wealth to satisfy the freehold qualification for running for representative to Georgia’s legislature. The estimate is similar in size to Bleakley and J. P. Ferrie’s (2013) estimate that winning a prize in the 1832 Georgia lottery significantly increases 1850 total census wealth (i.e., combined slave and real-estate wealth) by \$200.⁴

This paper proceeds as follows: Section 3.2 provides a theoretical discussion of the chan-

²The linear model used to obtain the estimate assumes that distance of the land to the city is uncorrelated with unobserved individual-level predictors of political labor supply, such as the opportunity costs associated with holding office.

³The six later Georgia lotteries were held in 1820, 1821, 1827, 1832, 1832 (“Gold Lottery”), and 1833 (“Fractions Lottery”).

⁴The authors use a sample of 1832 lottery participants linked to the full-count 1850 Census to investigate the effects of lottery prize values on the long-run wealth distribution of lottery winners. Using the same linked sample, Bleakley and J. Ferrie (2016) finds no evidence of a treatment effect on the wealth, literacy, or occupational standing of lottery winners or their descendants.

nels through which wealth may or may not impact political power; Section 3.3 provides some historical background, including details on the development and implementation of the 1805 and 1807 lotteries. Section 3.4 describes the data and record-linkage procedure, and provides summary statistics on the wealth of officeholders. Section 3.5 specifies the treatment effect estimator, tests the assumption of random treatment assignment, and provides treatment effect estimates. Section 3.6 discusses concerns regarding statistical power and the importance of treatment; Section 3.7 concludes.

3.2 Theoretical considerations

Through which channel should we expect wealth to influence officeholding, both in general and in the case of the antebellum South? In general, wealth makes it easier to access politics or makes it more important to have political power in order to protect vested interests. Materialist interpretations of oligarchies view personal wealth as a source of political power because it is tightly concentrated among a minority of the population, can be translated into political influence, and accompanies a set of political interests, such as wealth defense (Winters and Page 2009). According to that interpretation, wealthy citizens do not need to hold office or explicitly coordinate their political efforts, but instead work toward similar policy goals through political investment or lobbying.

In the case of antebellum Georgia, wealth shocks could have enabled poor citizens to overcome freehold requirements for holding office and attenuate the opportunity costs associated with candidacy and officeholding since the state legislature was not yet professionalized. Citizen-candidate models (Osborne and Slivinski 1996; Besley and Coate 1997) demonstrate that the number of candidates entering political contests declines with the cost of running, which is modeled both as campaigning cost and the opportunity cost of entering political races. In such models, the set of eligible candidates is the universe of eligible voters. In antebellum Georgia, the former was a proper subset of the latter because many otherwise eligible citizens could not run for office owing to freehold requirements for officeholding.

Corvalan, Querubin, and Vicente (forthcoming, 2018) build on the citizen-candidate model by adding a wealth-based eligibility requirements, which results in ideological distributions differing between eligible candidates and voters. Their model predicts that when the eligibility requirement is high enough to be binding — i.e., when the eligibility requirement exceeds the median wealth of the constituency — the key player is the citizen closest to the median within the set of eligible candidates. When the eligibility requirement does not bind, the key player is the constituency median. Eligibility requirements that are not binding can still restrict electoral competition in equilibrium by ruling out the possibility of poorer candidates who prefer higher taxes.

As described in further detail in Section 3.3, wealth-based eligibility requirements in Georgia were not binding in the sense that they did not exceed the estimated median property wealth at the time of the land lotteries. The lotteries initiated an exogenous shock to the wealth distribution that would have provided poor citizens with the freehold eligibility

required to run for the state legislature. In the context of Corvalan, Querubin, and Vicente (forthcoming, 2018), the exogenous wealth shock resulting from the land lotteries therefore had an effect on tax policy and electoral competition identical to lessening eligibility restrictions for officeholding: the decisive citizen becomes less wealthy and more favorable to income redistribution, potentially opening the door for “leftist” candidates.

Wealth shocks resulting from the lotteries likewise may have lowered the opportunity costs associated with officeholding, especially considering that Georgia’s legislature was not professionalized at the time. State legislatures that are not professionalized tend to be dominated by proprietors rather than wage and salary earners, who face higher opportunity costs of serving (Fiorina 1994, 1999). Studies of legislative careers and strategic retirements generally find that nonpolitical wealth and tenure in office are inversely related. Hall and Van Houweling (1995), for instance, find that a significant number of US Representatives decided to postpone retirement in 1990 in order to take advantage of a pension windfall enacted by the outgoing Congress. Groseclose and Krehbiel (1994) estimates that the opportunity cost of remaining in office accounts for more than a third of retirements in 1992, which was the final year in which Members of Congress could transform campaign funds into personal wealth.⁵

3.3 Historical background

Inequality in political representation was prevalent throughout the antebellum South. For example, Helper (1860) rails against the “slave-driving oligarchy” which dominated political offices then:

The magistrates in the villages, the constables in the districts, the commissioners of the towns, the mayors of the cities, the sheriffs of the counties, the judges of the various courts, the members of the legislatures, the governors of the States, the representatives and senators in Congress — are all slaveholders....There is no legislation except for the benefit of slavery and slaveholders.

The slaveholders’ dilemma was to secure political dominance amid universal white male suffrage. The 1798 Georgia constitution, which was in effect for the entirety of the antebellum period, extended rights to elect members of the legislature to adult white males. Propertyless whites were discouraged from voting on account of a poll tax that the legislature maintained during the antebellum period. Slaveholder candidates, however, often paid the poll taxes of poor whites in exchange for political support (Meyers and Williams 2012). Slaveholding cotton planters found allies in members of the clerical and the professional classes, who often owned a few slaves for personal service, and that alliance created a class of “social retainers” who defended the interests of slaveholding planters (Simons 1912). Echoing the

⁵Relatedly, Milyo and Groseclose (1999) find that incumbent wealth is orthogonal to electoral success in the 1992 House elections.

US Constitution of 1787, the three-fifths clause in the 1798 Georgia constitution counted slaves as three-fifths of a person for the purpose of representation in the legislature and thus fortified the slaveholders' control of the legislature (Coulter 1960).

Barriers to officeholding

The 1798 Georgia constitution also tied officeholding eligibility to land and property values. Beyond age and residency requirements, candidates for state representative were required to own freehold estates worth \$250 or possess \$500 of taxable property within their constituencies. The eligibility requirements for state senator were double those amounts. Qualifications for governor, an office opened to direct elections in 1824, included owning at least 500 acres of land within the state and other property to the value of \$4,000. Amendments to the constitution removed freehold qualifications for the legislature in 1835 and for governor in 1847.

In the terminology of Corvalan, Querubin, and Vicente (forthcoming, 2018), the freehold requirements for serving in the legislature were not binding because the requirements did not exceed the estimated median property wealth at the time of the land lotteries. Drawing a sample from the full-count 1850 Census (Ruggles et al. 2015) of adult male heads-of-household who were born in Georgia, living in Georgia, and had nonmissing surnames and property values, I estimate that about 13% of otherwise-eligible citizens had nominal property values below the freehold requirement for state representative and 28% were barred from running for state senator.

However, winning a land lottery prize enabled otherwise eligible lottery winners who did not satisfy freehold qualifications to hold office in the legislature. Using county-level data from the 1850 Census (Haines 2004), I estimate that the mean value of a land prize in either of the first two lotteries exceeded \$800, which represents 55% of median property wealth at the time of the lotteries.⁶

Georgia's legislature was not professionalized at the time of the lotteries and was mandated by the constitution to meet only once a year.⁷ The 1798 constitution provides salaries only for members of the executive branch and judiciary, while *per diem* compensation for legislators was not introduced until the 1877 constitution. The lack of legislative professionalism may have imposed an additional barrier for less wealthy citizens to run for the legislature owing to the opportunity costs associated with holding office.

Officeholder wealth

How does the wealth of Georgia's political class compare with the wealth of the general population? Following the record linkage procedure described in Section 3.4, I link the full

⁶Table B.1 provides information on the estimated lot value per county and a description of how the mean land prize values are calculated.

⁷The legislature typically met more than once. For example, in 1805 the House and Senate each met about 30 times.

list of officeholders to the 1820 and 1850 censuses in order to get a sense of where officeholders stand in the distribution of slave wealth and real estate wealth, respectively.⁸ I successfully link 39% and 66% of officeholders to the 1820 and 1850 Censuses, respectively. In comparison, Corvalan, Querubin, and Vicente (forthcoming, 2018) report success rates of between 41% and 68% when linking the names of state senators from the Carolinas and Virginia to the full-count 1850 Census.

Table B.2 reports basic summary statistics on the wealth-holding of officeholders versus the rest of the sample. Officeholders matched to the 1820 Census hold \$548 more in slave wealth, on average, than the rest of the sample ($p < 0.001$). Similarly, officeholders matched to the 1850 Census hold about \$400 more in real estate wealth, on average, than non-officeholders ($p < 0.001$).

The 1805 and 1807 lotteries

In the wake of public land fraud scandals, Georgia's legislature created a lottery system to distribute 1.3 million acres of newly acquired public land ceded by the Creek tribe. The 1805 lottery carved out three new counties from that land: Baldwin and Wilkinson counties, each divided into five districts, and Wayne county, divided into three districts (Clayton and Adams 1812). The 1807 lottery extended the boundary between Baldwin and Wilkinson, more than tripling the number of districts within the two counties.

Free white adult men and orphaned children were eligible for a single draw, while married men with children and widows with children were eligible for two draws in the 1805 lottery. The eligibility rules for the 1807 lottery were similar, with the main exception being that adult unmarried females could participate and 1805 lottery winners were excluded from participation.⁹ Registration for each lottery was voluntary and required a payment of 12.5 cents per draw. I estimate that approximately 85% of eligible males living in Georgia participated in the 1805 lottery and 70% of eligible males participated in the 1807 lottery.¹⁰

Prior to each lottery, a surveyor was appointed by the legislature to map the districts into square lots. During the 1805 lottery, tickets representing each lot were placed in a wooden lottery wheel to be awarded as prizes, along with blank tickets equal in number to the difference between the number of prizes and the number of draws. Two lottery wheels were used in the 1807 drawing: one containing the names of participants, and the other containing lot numbers. Blank tickets in number equal to difference between the number of

⁸The match probability averages are close to zero because the model is trained on a different domain; i.e., linking 1807 participants to officeholder records. I drop the prediction threshold to 25% in order to compensate for lower average match probabilities.

⁹Participants were required to take an oath when “doubt exist in the minds of the said justices” regarding the veracity of participants’ eligibility. The legislature criminalized making false statements concerning eligibility in the land lotteries. The law specifies that if found guilty in a jury trial, half of the defendant’s land went to the informer and the other half is reverted to the state, to be auctioned as land fractions.

¹⁰The 1805 calculation was made by taking the proportion of adult male participants over the total white male population aged 16 and over in the 1800 Census. The 1807 calculation subtracts the number of adult male winners in the 1805 lottery from the numerator.

registered draws and lot prizes were added to the lottery wheel. Tickets from each wheel were drawn simultaneously to form a combined ticket, and the participant won if the combined ticket included a prize. Lottery winners were required to pay \$4 per hundred acres for lots won in the 1805 lottery, or \$6 per hundred acres for lots won in the 1807 lottery, in order to obtain the title on the lot.¹¹

Winning a prize in the lottery represents a pure wealth shock because no homesteading requirement was imposed and lottery winners could easily sell their grants in a secondary market for public land (Weiman 1991). In some cases, lottery winners sold their grants to land speculators, who in turn sold the land to out-of-state settlers (Davis 1981). Land speculators often sought out lottery winners who drew particularly valuable lots (Cadle 1991).

3.4 Data and descriptive statistics

The primary source of data for this study is Graham's (2004) record of 1805 lottery winners and losers. The 1805 lottery is the only Georgia lottery to have recorded the names of all lottery participants. The records contain information on participants' names, county of registration, lottery draws, and lottery prizes won. Identifying remarks next to the participants' names provide additional information, such as generational suffix (i.e., "Jr." and "Sr.") and orphan status. Table 3.1 reports that 15% of the unrestricted sample of 1805 lottery participants ($N = 23,927$) received at least one lottery prize and was registered for 1.65 draws, on average, and 4% of 1805 lottery winners ($N = 3,707$) won two prizes.

I use lottery winner grant records for the 1807 lottery (Graham 2011) to construct a sample of 1807 lottery winners ($N = 8,822$), in which 14% of the sample won two prizes. Since the 1807 lottery records do not include information on the number of registered draws, I impute the number of draws by assigning two draws to all participants, except for women and orphans, who are assigned one draw. The mean of the imputed number of draws among 1807 winners, 1.85 draws, is virtually identical to the mean of the actual number of registered draws registered by 1805 winners.

In order to test the hypothesis that wealth increases the probability of officeholding, I link participant names to a historical roster of officeholders published in the *Georgia Official and Statistical Register* (Archives 1978, 1990). The roster includes information on names, jurisdictions, and term dates for all elected and appointed officeholders from the state's colonial period to 1990. I consider officeholders whose first term began in 1805 or 1807 to 1850, inclusive. That decision is based on the fact that the white male life expectancy at age 20 in the early 1800s was approximately 40 years (Hacker 2010), the youngest participants who were both eligible for the lottery and public office were 21 years of age and were expected to live 40 additional years.

¹¹Lottery winners had 12 months following the drawing to claim their prize under law; however, the legislature extended the deadline for claiming prizes on an annual basis until 1815. If lottery winners did not claim their land prior to the deadline, then the lots reverted to the state and sold in a public auction.

I employ a machine learning approach for linking lottery participants to officeholder records. First, I link 1807 lottery winners to officeholders based on an exact match on surname and a phonetic algorithm (i.e., Soundex) code of first name, and then deduplicate the matched records manually. Second, I train an ensemble of algorithmic models on the 1807 records to classify correct matches, using participant characteristics (e.g., the frequency and lengths of surnames) and match characteristics (e.g., the Euclidean distance between participants' county of registration and officeholders' constituency) as features of the model.¹² Lastly, I use the ensemble fit to deduplicate automatically 1805 participant records matched with officeholders on the basis of an exact match on surname and Soundex code of the first name; a prediction threshold of 50% is adopted to classify correct matches. Thus, the *Officeholder* outcome variable is a binary variable indicating whether the participant is linked to the officeholder records.

As reported in Table 3.1, 17% of adult males who participated in the 1805 lottery are matched to officeholder records successfully, and about one-in-ten of matched participants held office before the drawing of the 1805 lottery. Three-quarters of participants linked to officeholder records first served in the state House, while the rest started their political careers in the state Senate, US House, or other state executive branch offices.

Candidacy

To address the question of whether wealth makes individuals more likely to be candidates for office, I extract candidate names from two election datasets, the first covering all offices from the local level to the federal level from 1787 to 1825 (Lampi 2013) and the second covering federal offices from 1788 to 1990 (ICPSR 1984). I link participants to candidates who first ran for office between the period of 1805 or 1807 to 1850 using the same deduplication model used for linking participants to officeholder names.¹³ The resulting *Candidate* outcome is a binary variable that captures whether participants ran for office. Since the candidate records are more limited in scope compared to the officeholder records, only about 2% of 1805 and 1807 lottery participants are successfully linked to candidate records, and about a quarter of these matched participants ran for office prior to 1805 or 1807.

Future wealth

Finally, I investigate whether winning a lottery prize increased long-term wealth by linking the participant records to the full-count 1820 Census and estimating the treatment effect on imputed slave wealth. The 1820 Census is the earliest surviving enumeration of Georgia's population and represents all counties except for Franklin, Rabun, and Twiggs counties. The records include information on the name of the head of household and the number of slaves

¹²The cross-validated mean squared error on the training set is less than 3%. Table B.3 provides information on the record link ensemble's candidate learners, weights, and error estimates.

¹³I lower the prediction threshold to 12.5% to adjust for candidate match probabilities that average near zero.

held by gender and age group, which I use to impute slave wealth.¹⁴ I match 18% of 1805 lottery participants and 40% of 1807 lottery winners to the 1820 Census. The mean slave wealth in the linked samples is between \$1,222 and \$1,971, which is approximates the market value of two male prime-age field hands in Georgia at the time of the lotteries.¹⁵

3.5 Estimation and results

I estimate the following linear model:

$$y_{i,s} = \# \text{ draws}_{i,s} + \delta \text{ treat}_{i,s} + \epsilon_{i,s}, \quad (3.1)$$

where $y_{i,s}$ is the candidate or officeholder response or 1820 slave wealth measure for participant $i = \{1, \dots, N\}$ in sample s and $\# \text{ draws}_{i,s} \in \{1, 2\}$ represents the number of draws registered by each participant. When the sample includes both 1805 winners and losers, $\text{treat}_{i,s}$ is a binary treatment assignment variable that assumes a value of unity for participants who won at least one prize in the 1805 lottery and assumes a zero value for participants who did not win a prize. The coefficient of interest, δ , corresponds to the intention-to-treat (ITT) estimate of the sample average treatment effect. When the sample is limited to winners in either the 1805 or 1807 lotteries, $\text{treat}_{i,s}$ is equal to unity for participants who win two prizes in the lottery and is zero otherwise. In that case, the ITT effect captures the marginal effect of winning a second land prize.

Balance

The model assumes that treatment assignment is random conditional on the number of registered draws. While that assumption cannot be tested directly, we can verify that treatment assignment with respect to the pretreatment covariates is balanced. Statistically significant treatment effects on pretreatment covariates at the level of $\alpha = 0.05$ indicate imbalance in treatment assignment.

Fig. 3.1 plots the p -values corresponding to the treatment effect estimated by estimating Eq. 3.1 on each pretreatment covariate summarized in Table 3.1. Treatment assignment is balanced across all pretreatment covariates for 1805 lottery winners and losers. When the sample is restricted to 1805 lottery winners, winners registered in Clarke County were more likely to have won two prizes rather than a single prize, controlling for the number of draws ($p = 0.02$). However, the difference is not significant when accounting for the multiple comparisons made for the balance tests.¹⁶

¹⁴Access to the full-count data is granted by agreement between UC Berkeley, and the Minnesota Population Center (Ruggles et al. 2015). The Minnesota Population Center has collected digitized census data for 1790-1930 microdata collection with contributions from Ancestry.com and FamilySearch.

¹⁵Phillips (1905) estimates the average value of male prime field hands (18-30 years old) in Georgia in 1821 is \$700. I impute slave wealth in 1820 by using the coefficients in Table II of Kotlikoff (1979) to adjust this average value according to age group and gender.

¹⁶The significance level for the Bonferroni correction is $\alpha = 0.05/33 = 0.001$.

Table 3.1: Distribution of pretreatment and response variables by sample.

Variable	Sample	N	Min.	Mean	Max.	S.d.
<i>Pretreatment variables</i>						
1820 Census Match	1805 winners & losers	23,927	0	0.18	0.94	0.34
	1805 winners	3,707	0	0.18	0.93	0.34
	1807 winners	8,822	0	0.4	0.94	0.15
Candidate		0	0.01	1	0.07	
		0	0.01	1	0.07	
		0	0.01	1	0.06	
Officeholder		0	0.02	1	0.15	
		0	0.02	1	0.15	
		0	0.02	1	0.15	
Junior		0	0.03	1	0.18	
		0	0.03	1	0.16	
		0	0.02	1	0.14	
Senior		0	0.03	1	0.18	
		0	0.04	1	0.2	
		0	0.03	1	0.17	
Surname Frequency		1	32.61	348	58.62	
		1	32.6	348	57.49	
		1	11.6	127	19.9	
Surname Length		3	6.26	14	1.59	
		3	6.28	13	1.59	
		1	6.31	15	1.66	
<i>Treatment variables</i>						
Treated	1805 winners & losers	23,927	0	0.15	1	0.36
	1805 winners	3,707	0	0.04	1	0.19
	1807 winners	8,822	0	0.14	1	0.34
# draws		1	1.65	2	0.48	
		1	1.82	2	0.38	
		1	1.8	2	0.4	
<i>Outcome variables</i>						
Candidate	1805 winners & losers	21,732	0	0.02	1	0.14
	1805 winners	3,346	0	0.02	1	0.13
	1807 winners	6,986	0	0.01	1	0.11
Officeholder		0	0.17	1	0.37	
		0	0.17	1	0.38	
		0	0.14	1	0.35	
Slave wealth (1820\$)	1805 winners & losers	5,252	0	1,222.62	31,227.84	2,234.66
	1805 winners	828	0	1,321.78	27,709.36	2,531.75
	1807 winners	8,534	0	1,971.64	29,882.27	845.93

Notes: *Treated* is defined as winning at least one prize for the 1805 winners and losers sample, and winning two prizes for the samples restricted to winners. *# draws* is the number of registered draws. *Candidate* and *Officeholder* indicates whether participants ran for office or held office, respectively. *Slave wealth* (in 1820 dollars) is the imputed slave wealth for participants matched to the 1820 Census.

In the sample of 1807 lottery winners, winners who ran for office prior to the 1807 lottery were more likely to have won a second prize ($p = 0.01$). Additionally, 1807 lottery winners were more likely to have won a second prize if they also were more likely to be matched to the 1820 Census ($p < 0.01$), have names with generational suffixes ($p < 0.01$), or have more common surnames ($p < 0.01$). Treatment assignment likewise is unbalanced for 1807 winners registered in the counties of Bryan, Clarke, and Lincoln.

ITT estimates

Table 3.2 presents the ITT treatment effect estimates on each outcome of interest. The confidence intervals for the ITT estimates on officeholding tightly straddle zero for each sample. In the discussion below, I argue that these null results on officeholding are informative because the estimated effects are not practically different from zero. For example, the upper bound of the interval of the treatment effect estimate on *Officeholder* implies that, at most, 1% of the sample treated group would select into office as a result of receiving treatment. The confidence interval for the estimate is unchanged when estimating the effect on officeholder match probability (Table B.4).

While I find no evidence of a treatment effect on candidacy for the sample of 1805 winners and losers, the point estimates for the samples of 1805 and 1807 winners imply that the marginal effect of winning a second prize in the lottery significantly increases the probability of running for office by 1% to 2%. However, those estimates are not robust to using candidate match probability as the outcome variable (Table B.5).

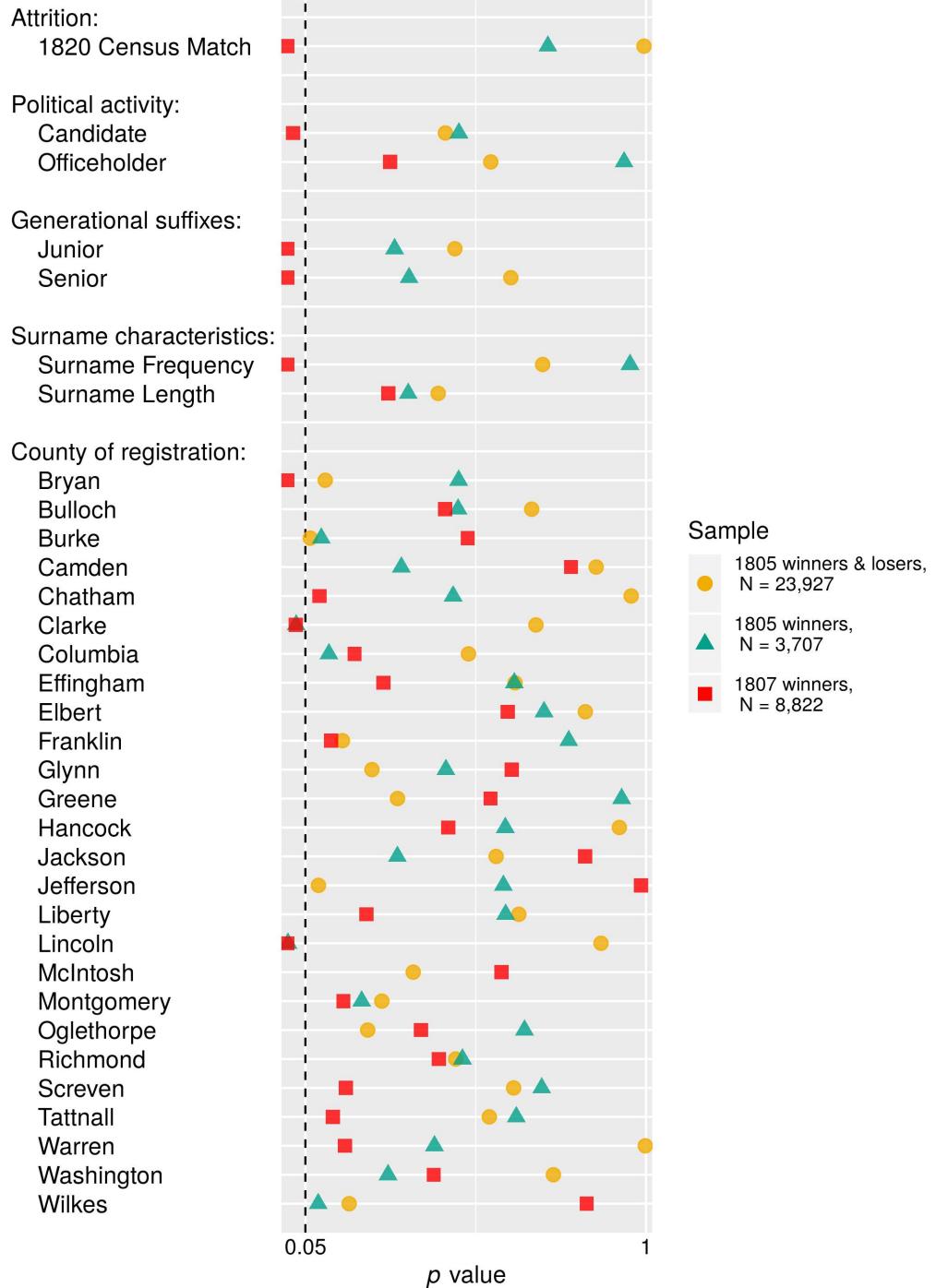
Lastly, I find no evidence that winning a prize in the 1805 lottery increased future slave wealth. In the sample of 1807 winners, winning a second prize significantly reduces slave wealth in 1820 by \$163 [-\$256, -\$69]; although, that estimate is not robust to including pretreatment covariates in the regression or weighting slave wealth by the probability of being matched to the 1820 Census (Table B.6).

Table 3.2: ITT treatment effect estimates.

Outcome Sample \ \backslash	Candidate	Officeholder	Slave wealth (1820\$)
1805 winners & losers	-0.001 [-0.007, 0.003]	0.003 [-0.01, 0.01]	156.68 [-51.13, 364.49]
1805 winners	0.02 [0.0004, 0.05]	-0.01 [-0.08, 0.05]	-155.12 [-1,251.37, 941.12]
1807 winners	0.008 [0.0008, 0.01]	-0.004 [-0.02, 0.01]	-163.08 [-256.34, -69.82]

Notes: values in brackets represent 95% confidence intervals for treatment effect estimates derived from the standard errors of the linear model (Eq. 3.1). Orphans and women are excluded from each sample for candidacy and officeholding outcomes.

Figure 3.1: Balance in treatment assignment by lottery sample.



Notes: *1820 Census Match* is the probability of being linked to the 1820 Census; *Candidate* and *Officeholder* indicates participants who ran for or held office, respectively, prior to the 1805 lottery. Samples include all lottery participants.

3.6 Discussion

In the short-run, treatment enabled lottery winners at the lower-end of the wealth distribution to meet the freehold qualifications for holding office in the state legislature. Still, treatment may be too weak to cause a substantively meaningful increase in the likelihood that lottery winners would select into public office, given how rare an event it is for citizens to become elected officials. I conduct a power analysis by simulation to ensure that the research design allows for the identification of a significant treatment effect. The simulation results imply that if the actual treatment effect size in a hypothetical finite population is 2.1%, the research design provides an 80% chance of rejecting the null hypothesis that treated and control participants are equally likely to hold office (Fig. B.1). Since I am unable to reject the null in the sample, it follows that the population effect size is most likely less than 2.1%, an effect not practically different from zero.

Is it the case that the null effects for officeholding are driven by the fact that lottery wealth does not translate into larger long-run wealth, but rather increases present consumption? Lottery winners may perceive lottery wealth as a financial windfall and spend the winnings more quickly than earned wealth (Doherty, Gerber, and Green 2006). While I find no evidence that winning a prize in the 1805 lottery increased the future wealth of lottery winners, I provide evidence from quantile regression estimates that treatment increased future wealth for 1805 lottery participants near the median of the wealth distribution (Fig. B.2). Specifically, I find that treatment confers a \$171 [\$18, \$323] increase in wealth for participants at the median. The upper bound of the confidence interval represents an increase in wealth sufficient to satisfy the freehold qualification for running for state representative.

While winning a land lottery prize enabled otherwise eligible lottery winners who did not satisfy freehold qualifications to hold office in the legislature, pervasive wealth inequality may have played a role in stifling access to politics in Georgia in the early 19th century. In 1850, the top decile of property-owners held over half of the total property wealth in that state.¹⁷ It may be difficult in this setting for even the *nouveau riche* to overcome barriers to entry into politics, such as the opportunity costs associated with participating in politics. Further investigations in alternative settings would help determine whether the results of the present study would be different in a context of less wealth inequality.

3.7 Conclusion

This study exploits the first two land lotteries in US history to estimate the effect of lottery wealth on subsequent political power. Personal wealth is expected to reduce the opportunity costs of holding office and may make it more important for the wealthy to seek and to hold

¹⁷The calculation uses a sample of adult male heads of household who were born in Georgia and were living in Georgia at the time of the 1850 Census. The statewide slave wealth Gini coefficient in 1820 is 0.78 and the coefficient for statewide real estate wealth in 1850 is 0.66, both indicating substantial wealth inequality.

office. In the case of antebellum Georgia, wealth shocks could have enabled poor citizens to overcome freehold officeholding requirements and lower the opportunity costs associated with officeholding since Georgia's legislature was not yet professionalized. I find no evidence in support of the hypotheses that wealth increases the probability of running for or holding office.

The null findings nevertheless are informative because the estimated effects are not practically different from zero. What is more important, the absence of a treatment effect implies that commonly observed cross-sectional correlations between personal wealth and public officeholding are likely explained by selection effects.

Chapter 4

RNN-Based Counterfactual Prediction, with an Application to Homestead Policy and Public Schooling

Summary: This paper proposes an alternative to the synthetic control method (SCM) for estimating the effect of a policy intervention on an outcome over time. Recurrent neural networks (RNNs) are used to predict the counterfactual outcomes of treated units using only the outcomes of control units as predictors. This approach is less susceptible to p -hacking because it does not require the researcher to choose predictors or pre-intervention covariates to construct the synthetic control. RNNs do not assume a functional form, can learn nonconvex combinations of control units, and are specifically structured to exploit temporal dependencies in sequential data. I apply the approach to the problem of estimating the long-run impacts of U.S. homestead policy on public school spending.

4.1 Introduction

An important problem in the social sciences is estimating the effect of a discrete intervention on a continuous outcome over time. When interventions take place at an aggregate level (e.g., a state), researchers make causal inferences by comparing the post-intervention (“post-period”) outcomes of affected (“treated”) units against the outcomes of unaffected (“control”) units. A common approach to the problem is the synthetic control method (SCM) (Abadie, Diamond, and Hainmueller 2010), which predicts the counterfactual outcomes of treated units by finding a convex combination of control units that match the treated units in term of lagged outcomes or pre-intervention (“pre-period”) covariates.

The SCM has several limitations. First, the convexity restriction of the synthetic control estimator precludes dynamic, nonlinear interactions between multiple control units. Intuitively, one can expect that the treated unit may exhibit nonlinear or negative correlations

with the control units. Ferman and Pinto (2016) demonstrate that the convexity restriction implies that the SCM estimator may be biased even if selection into treatment is only correlated with time-invariant unobserved covariates. Second, Ferman and Pinto (2018) demonstrate that the SCM is generally biased if treatment assignment is correlated with unobserved confounders, even when the number of pre-period periods grows. Moreover, the authors show that while the SCM minimizes imbalance in pre-period outcomes, the likelihood of finding exact balancing weights vanishes as the number of time periods increase, which results in bias.

While the strength of the SCM lies in its simplicity in setup and implementation, several problems arise from the lack of guidance on how to specify the SCM estimator. The specification of the estimator can produce very different results: Ferman, Pinto, and Possebom (2018) show, for example, how cherry-picking between common SCM specifications can facilitate *p*-hacking. Kaul et al. (2015) show that the common practice of including lagged outcomes as model inputs can render all other covariates irrelevant. Lastly, Klößner et al. (2017) demonstrates that the common practice of using cross-validation to select importance weights can yield multiple values and consequently different results.

This paper proposes an alternative to the SCM that is capable of automatically selecting appropriate control units at each time-step, allows for nonconvex combinations of control units, and does not rely on pre-period covariates. The method uses recurrent neural networks (RNNs) to predict the counterfactual outcomes of treated units using only control unit outcomes as model inputs. RNNs are a class of neural networks that take advantage of the sequential nature of temporal data by sharing model parameters across multiple time-steps (El Hihi and Bengio 1995). RNNs are nonparametric in that they do not assume a functional form when fitting the data. In addition, RNNs can learn the most useful nonconvex combination of control unit outcomes at each time-step for generating counterfactual predictions. Relaxing the convexity restriction is useful when the data-generating process underlying the outcome of interest depends nonlinearly on the history of its inputs. RNNs have been shown to outperform various linear models on time-series prediction tasks (Cinar et al. 2017).

RNNs are end-to-end trainable and very flexible to a given sequential prediction problem. For example, they are capable of sharing learned parameters across time-steps and multiple treated units. While the SCM can be generalized to handle multiple treated units (e.g., Dube and Zipperer 2015; Xu 2017), the generalized the SCM is not capable of sharing model weights when predicting the outcomes of multiple treated units. Regularization methods such as dropout can easily be incorporated into RNN architectures to prevent overfitting during the training process, which is problematic when the networks learn an overreliance on a few model inputs.

The proposed method builds on a new literature that uses machine learning methods for data-driven counterfactual prediction, such as matrix completion (Athey et al. 2017), or two-stage estimators that reduce data dimensionality via L1-regularized regression (Doudchenko and Imbens 2016; Carvalho, Masini, and Medeiros 2018) or matrix factorization (Amjad, Shah, and Shen 2018) prior to regressing the outcomes on the reduced data. These methods are data-driven in the sense that they are capable of finding an appropriate subset of control

units for comparison in the absence of domain knowledge or pre-period covariates.

In the section immediately below, I describe the problem of counterfactual prediction and its relationship to matrix completion and the problem of covariate shift; Section 4.3 introduces the approach of using RNNs for counterfactual prediction; Section 4.4 presents the results of the placebo tests; Section 4.5 details the procedure for hypothesis testing and applies the RNN-based method and inferential procedure to the problem of estimating the impact of homestead policy on long-run state government investment in public schooling; Section 4.6 concludes and offers potential avenues for future research.

4.2 Counterfactual prediction

The proposed method estimates the causal effect of a discrete intervention in observational panel data; i.e., settings in which treatment is not randomly assigned and there exists both pre- and post-period observations of the outcome of interest. Let \mathbf{Y} denote a $N \times T$ matrix of outcomes for each unit $i = 1, \dots, N$, at time $t = 1, \dots, T$. \mathbf{Y} is incomplete because we observe each element Y_{it} for only the control units and the treated units prior to time of initial treatment exposure, $T_0 < T$. Let \mathcal{O} denote the set of (it) values that are observed and \mathcal{M} the set of (it) missing values. Let the values of the $N \times T$ complete matrix \mathbf{W} be $W_{it} = 1$ if $(it) \in \mathcal{M}$ and $W_{it} = 0$ if $(it) \in \mathcal{O}$. The pattern of missing data is assumed throughout this paper to follow a simultaneous treatment adoption setting, where treated units are exposed to treatment at time T_0 and every subsequent period.

This setup is motivated by the Neyman (1923) potential outcomes framework, where for each it value there exists a pair of potential outcomes, $Y_{it}(1)$ and $Y_{it}(0)$, representing the response to treated and control regimes, respectively. The observed outcomes are

$$Y_{it} = \begin{cases} Y_{it}(0) & \text{if } W_{it} = 0 \text{ or } t < T_0 \\ Y_{it}(1) & \text{if } W_{it} = 1 \text{ and } t \geq T_0. \end{cases}$$

The problem of counterfactual prediction is that we cannot directly observe the missing potential outcomes and instead wish to impute the missing values in $\mathbf{Y}(0)$ for treated units with $W_{it} = 1$. The potential outcomes framework explicitly assumes unconfoundedness. In an observational setting, this assumption requires

$$(\mathbf{Y}(0), \mathbf{Y}(1)) \perp\!\!\!\perp \mathbf{W} | \mathbf{Y}(\mathcal{O}),$$

where $\mathbf{Y}(\mathcal{O})$ is the observed data.

The potential outcomes framework also implicitly assumes treatment is well-defined to ensure that each unit has the same number of potential outcomes (Imbens and Rubin 2015). It also excludes interference between units, which would undermine the framework by creating more than two potential outcomes per unit, depending on the treatment status of other units (Rubin 1990).

Relationship to matrix completion and covariate shift

The intuition behind the proposed approach to counterfactual prediction is similar to that of the method of matrix completion via nuclear norm minimization (MC-NNM) proposed by Athey et al. (2017). Matrix completion methods attempt to impute missing entries in a low-rank matrix by solving a convex optimization problem via NNM, even when relatively few values are observed in \mathbf{Y} (Candès and Recht 2009; Candes and Plan 2010). The estimator recovers a $N \times T$ low-rank matrix by minimizing the sum of squared errors via nuclear norm regularized least squares. The estimator reconstructs the matrix by iteratively replacing missing values with those recovered from a singular value decomposition (Mazumder, Hastie, and Tibshirani 2010).

Athey et al. (2017) note two drawbacks of MC-NNM. First, the errors may be autocorrelated because the estimator does not account for temporal dependencies in the observed data. The estimator estimate patterns row- and column-wise, but treat the data as perfectly synchronized (Yoon, Zame, and Schaar 2018). In contrast, the SCM assumes that correlations across units are stable over time, while the RNN-based approach exploits the temporal component of the data and therefore does not have the problem of autocorrelated errors.

Second, the MC-NNM estimator penalizes the errors for each observed value equally without regard to the fact that the probability of missingness (i.e, the propensity score), increases with t . Athey et al. (2017) suggest weighting the loss function by the propensity score, which is similar to the importance weighting scheme proposed by Cortes et al. (2008) to address the problem of covariate shift, which is a special case of domain adaptation (Huang et al. 2007; Ben-David et al. 2007; Bickel, Brückner, and Scheffer 2009; Cortes, Mansour, and Mohri 2010; Ganin et al. 2015). Schnabel et al. (2016) first connected the matrix completion problem with causal inference in observational settings in the context of recommender systems under confounding. Johansson, Shalit, and Sontag (2016) formulates the general problem of counterfactual inference as a covariate shift problem.

The covariate shift problem occurs when training and test data are drawn from different distributions. For notational ease, define the training set input-output pair as

$$(\mathbf{X}^{\text{train}}, \mathbf{Y}^{\text{train}}) = (\mathbf{Y}(\mathbf{W})^{(t < T_0)}, \mathbf{Y}(\mathbf{W})^{(t \geq T_0)})$$

for units with $\mathbf{W} = 0$ and the test set pair $(\mathbf{X}^{\text{test}}, \mathbf{Y}^{\text{test}})$ for units with $\mathbf{W} = 1$. In the proposed approach, the model weights learned on the training set is fit on \mathbf{X}^{test} to predict \mathbf{Y}^{test} . The approach therefore assumes similarity between the distributions of $\mathbf{X}^{\text{train}}$ and \mathbf{X}^{test} . In order to minimize the discrepancy between the training and test set input distributions, I estimate the propensity score $\hat{e}_{it} = \Pr(W_{it} = 1 | Z_{it})$, conditional on covariate matrix \mathbf{Z} and then weight the training loss by the estimated propensity scores.

Nonparametric regression

In its most basic form, counterfactual prediction can be represented as a nonparametric regression of the training set outputs on the inputs,

$$\hat{\mathbf{Y}}^{\text{train}} = \hat{f}_0(\mathbf{X}^{\text{train}}) + \epsilon^{(t)}, \quad (4.1)$$

where the noise variables $\epsilon^{(t)}$ are assumed to be i.i.d. standard normal and independent of the observed data. The nonlinear function \hat{f}_0 is estimated by minimizing the weighted mean squared error on the training set outputs,

$$\text{WMSE} = \sum \left(\mathbf{Y}^{\text{train}} - \hat{\mathbf{Y}}^{\text{train}} \right)^2 \cdot \frac{\hat{\mathbf{E}}^{\text{train}}}{|\mathbf{X}^{\text{train}}|}, \quad (4.2)$$

where $\hat{\mathbf{E}}^{\text{train}}$ is a matrix of estimated propensity scores.

At test time, the estimated function is used to predict $\hat{\mathbf{Y}}^{\text{test}} = \hat{f}_0(\mathbf{X}^{\text{test}})$. The estimated causal effect of the intervention is then

$$\hat{\boldsymbol{\Phi}} = \mathbf{Y}^{\text{test}} - \hat{\mathbf{Y}}^{\text{test}}. \quad (4.3)$$

The estimated average causal effect of the intervention on treated units is calculated by averaging over the time dimension, resulting in the vector $\bar{\boldsymbol{\Phi}}^{(t)}$ of length $T_\star = T - T_0$.

4.3 RNNs for counterfactual prediction

RNNs (Graves 2012; Goodfellow, Bengio, and Courville 2016) consist of an input $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n_x)})$, an output $\mathbf{Y} = (\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n_y)})$, and a hidden state $\mathbf{h}^{(t)}$. In the plain vanilla RNN it is assumed $n_x = n_y = T$; in the encoder-decoder network architecture described below, n_x and n_y can vary in length.

At each t , RNNs input $\mathbf{x}^{(t)}$ and pass it to the $\mathbf{h}^{(t)}$, which is updated with a function $g^{(t)}$ using the entire history of the input, which is unfolded backwards in time:

$$\mathbf{h}^{(t)} = g^{(t)}\left(\mathbf{x}^{(t)}, \mathbf{x}^{(t-1)}, \dots, \mathbf{x}^{(1)}\right) \quad (4.4)$$

$$= f_1\left(\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}; \theta\right). \quad (4.5)$$

The activation function $f_1(\cdot)$, parameterized by θ , is shared for all t . Parameter sharing is particularly useful in the current application because it allows for better generalization when the dimension of the training data is relatively small. The updated hidden state (4.5) is used to generate a sequence of values $\mathbf{o}^{(t)}$ in the form of log probabilities corresponding to the output. The loss function computes $\hat{\mathbf{y}}^{(t)} = f_2(\mathbf{o}^{(t)})$ and calculates the loss. The total loss for the input-output pair is the sum of the losses over all t .

The RNNs are trained to estimate the conditional distribution of $\mathbf{y}^{(t)}$ given the past inputs and also the previous output. This is accomplished by offsetting the input-output pairs by one time-step so that the networks receive $\mathbf{y}^{(1)}$ as input at $t + 1$ to be conditioned on for predicting subsequent outputs. This popular training procedure is known as teacher forcing because it forces the networks to stay close to the ground-truth output $\mathbf{y}^{(t)}$ (Lamb et al. 2016). Specifically, the RNNs are trained to maximize the log-likelihood

$$\log \Pr\left(\mathbf{y}^{(t)} | \mathbf{x}^{(1)} \dots \mathbf{x}^{(t)}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(t-1)}\right). \quad (4.6)$$

Encoder-decoder networks

Encoder-decoder networks are the standard for neural machine translation (NMT) (Cho et al. 2014; Bahdanau, Cho, and Bengio 2014; Vinyals et al. 2014) and are also widely used for predictive tasks, including speech recognition (Chorowski et al. 2015) and time-series forecasting (Zhu and Laptev 2017).

The encoder RNN reads in $\mathbf{x}^{(t)}$ sequentially and the hidden state of the network updates according to (4.5). The hidden state of the encoder is a context vector \mathbf{c} that summarizes the input sequence, which is copied over to the decoder RNN. The decoder generates a variable-length output sequence by predicting $\mathbf{y}^{(t)}$ given the encoder hidden state and the previous element of the output sequence. Thus, the hidden state of the decoder is updated recursively by

$$\mathbf{h}^{(t)} = f_1\left(\mathbf{h}^{(t-1)}, \mathbf{y}^{(t-1)}, \mathbf{c}; \theta\right), \quad (4.7)$$

and the conditional probability of the next element of the sequence is

$$\Pr(\mathbf{y}^{(t)} | \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(t-1)}, \mathbf{c}) = f_1\left(\mathbf{h}^{(t)}, \mathbf{y}^{(t-1)}, \mathbf{c}; \theta\right). \quad (4.8)$$

Effectively, the decoder learns to generate outputs $\mathbf{y}^{(t)}$ given the previous outputs, conditioned on the input sequence.

Recurrent variational autoencoder

While the encoder-decoder architecture is effective for many sequential prediction tasks, the model does not learn a vector representation of the entire input. The variational autoencoder (VAE) (Kingma and Welling 2013) is a generative model that learns a latent variable model for $\mathbf{x}^{(t)}$ such that new sequences $\mathbf{x}'^{(t)}$ can be generated by sampling from the latent space q . Similar to encoder-decoder networks, the VAE has an encoder that learns a latent representation of the input sequence and a decoder that maps the representation back to the inputs. The VAE architecture differs from encoder-decoder networks in that the VAE doesn't have a final dense layer that compares the decoder outputs to $\mathbf{x}'^{(t)}$; i.e., it is a “self-supervised” technique. Another difference is that the VAE learns parameter weights by mapping the inputs to a distribution over parameters of q .

The recurrent VAE (RVAE) (Fabius and Amersfoort 2014; Chung et al. 2015; Bowman et al. 2015) consists of an encoder RNN that maps $\mathbf{x}^{(t)}$ to a distribution over parameters of q . The model then randomly samples \mathbf{z} from the latent distribution,

$$q(\mathbf{z}|\mathbf{x}^{(t)}) = q(\mathbf{z}; f_3(\mathbf{x}^{(t)}; \theta)). \quad (4.9)$$

Finally, a decoder RNN takes the form of a conditional probability model $\text{Pr}(\mathbf{x}^{(t)}|\mathbf{z})$. The parameters of the model are learned by maximizing the loss function, which takes the difference between the log-likelihood between the decoder outputs $\mathbf{x}'^{(t)}$ and $\mathbf{x}^{(t)}$ and the relative entropy between $q(\mathbf{z}|\mathbf{x}^{(t)})$ and the model prior $\text{Pr}(\mathbf{z})$. The latter component of the loss function acts as regularizer by forcing the learned latent distribution to be similar to the model prior.

4.4 Placebo tests

I conduct placebo tests on actual datasets in order to benchmark the accuracy of RNN-based estimators. There are no actual treated units in the placebo tests, so the estimators are evaluated on their ability to recover a null effect.

For each trial run, I randomly select half of the units in the dataset to be treated and predict their counterfactual outcomes for periods following a selected T_0 . I compare the predicted values to the observed values by calculating the root-mean squared error (RMSE). I benchmark the encoder-decoder networks and RVAE against the following estimators:

- (a) **DID** Regression of \mathbf{Y} on \mathbf{W} and unit and time fixed effects
- (b) **MC-NNM** Matrix completion via nuclear norm minimization, with the regularization term on the nuclear norm selected by cross-validation (Athey et al. 2017)
- (c) **SCM** Approached via exponentiated gradient descent (Abadie, Diamond, and Hainmueller 2010)
- (d) **VT-EN** Vertical regression with elastic-net regularization, with the regularization and mixing parameters selected by cross-validation (Zou and Hastie 2005; Athey et al. 2017).

Implementation details for the encoder-decoder networks and RVAE are provided in Supporting Materials (SM) Section SM-E. In the placebo tests, the networks are trained using an unweighted MSE loss function for 500 epochs on a 12GB NVIDIA Titan Xp GPU.

Synthetic control datasets

I first conduct placebo tests on three datasets common to the synthetic control literature, with the actual treated unit removed from each dataset: Abadie and Gardeazabal's (2003)

study of the economic impact of terrorism in the Basque Country during the late 1960s ($N = 16$, $T = 43$); Abadie, Diamond, and Hainmueller's (2010) study of the effects of a large-scale tobacco control program implemented in California in 1988 ($N = 38$, $T = 31$); and Abadie, Diamond, and Hainmueller's (2015) study of the economic impact of the 1990 German reunification on West Germany ($N = 16$, $T = 44$). Each dataset is log-transformed to alleviate exponential effects.

Figure 4.1 reports the estimated average prediction error on the California smoking dataset, with the estimates jittered horizontally to reduce overlap. Figures SM-D.1 and SM-D.2 report the estimates for the Basque Country and West Germany datasets, respectively. Error bars are calculated using the standard deviation of the error distribution generated by multiple runs. The RNN-based estimators yield comparable error rates vis-à-vis the alternatives only for high ratios of T_0/T , which reflect the need for sizeable training sets for the RNN-based approach. The RVAE performs the worse on comparatively small training data since it is learning from less information than the encoder-decoder networks; i.e., without the post-period observations of the control units. The MC-NNM estimator does comparatively well in the simulations due to the fact that it is capable of using additional information in the form of pre-period observations of the treated units, whereas the other estimators train only on the control observations.

Stock market data

The second battery of placebo tests draws on a dataset of stock market returns compiled by Athey et al. (2017). The dataset consists of daily returns for 2,453 stocks over 3,082 days. In order to track how the error rates vary according to the dimensionality of the data, I create six sub-samples of the first T daily returns of N randomly selected stocks for the pairs $(N, T) = (10, 490)$, $(20, 245)$, $(50, 98)$, $(70, 70)$, $(100, 49)$, and $(140, 35)$. In each sub-sample, half of the units are randomly selected as treated, and $T_0 = T/2$.

Figure 4.2 reports the average RMSE for each pair with standard errors informed by the error distribution generated by five trial runs. The average RMSE is the lowest for all estimators in the sub-sample $(N, T) = (10, 490)$, which reflects the benefit of training on a large number of time periods. Within this sub-sample, encoder-decoder networks and RVAE achieve the lowest average RMSE, followed by MC-NNM, SCM, DID, and lastly, vertical regression. The RNN-based estimators do comparatively less well when $N \gg T$ since there is not an adequate number of training set pre-periods to learn a concise representation of the inputs.

4.5 Application: Homestead policy and public schooling

Sociologists and political economists (e.g, Meyer et al. 1979; Alesina and Reich 2013; Bandiera et al. 2018) have viewed the rapid development of public schooling in the U.S. during the 19th century as a nation-building policy. It is argued that states across the U.S. adopted

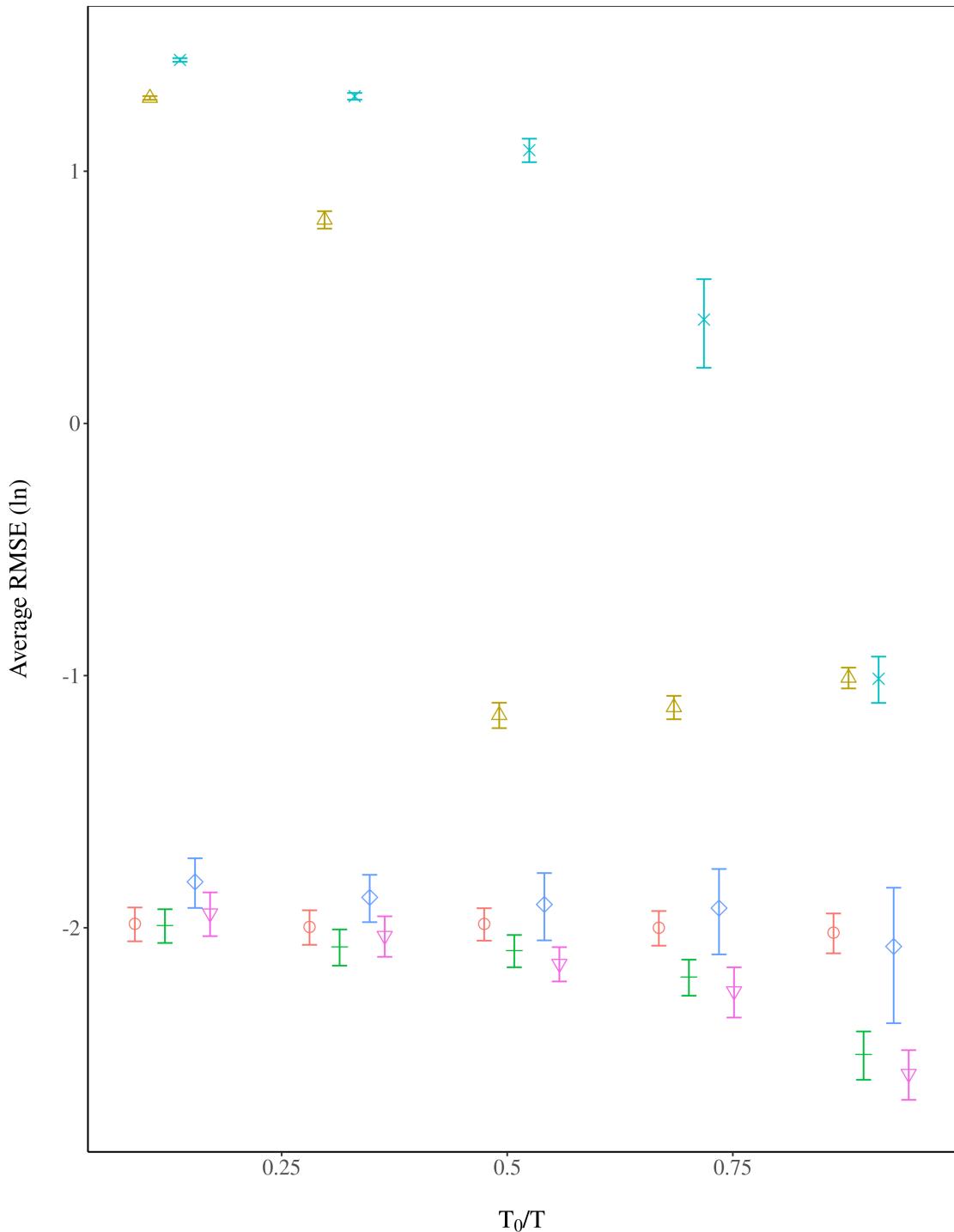


Figure 4.1: Placebo tests on California smoking data: \ominus , DID; \triangle , ED; $+$, MC-NNM; \times , RVAE; \diamond , SCM; ∇ , VT-EN.

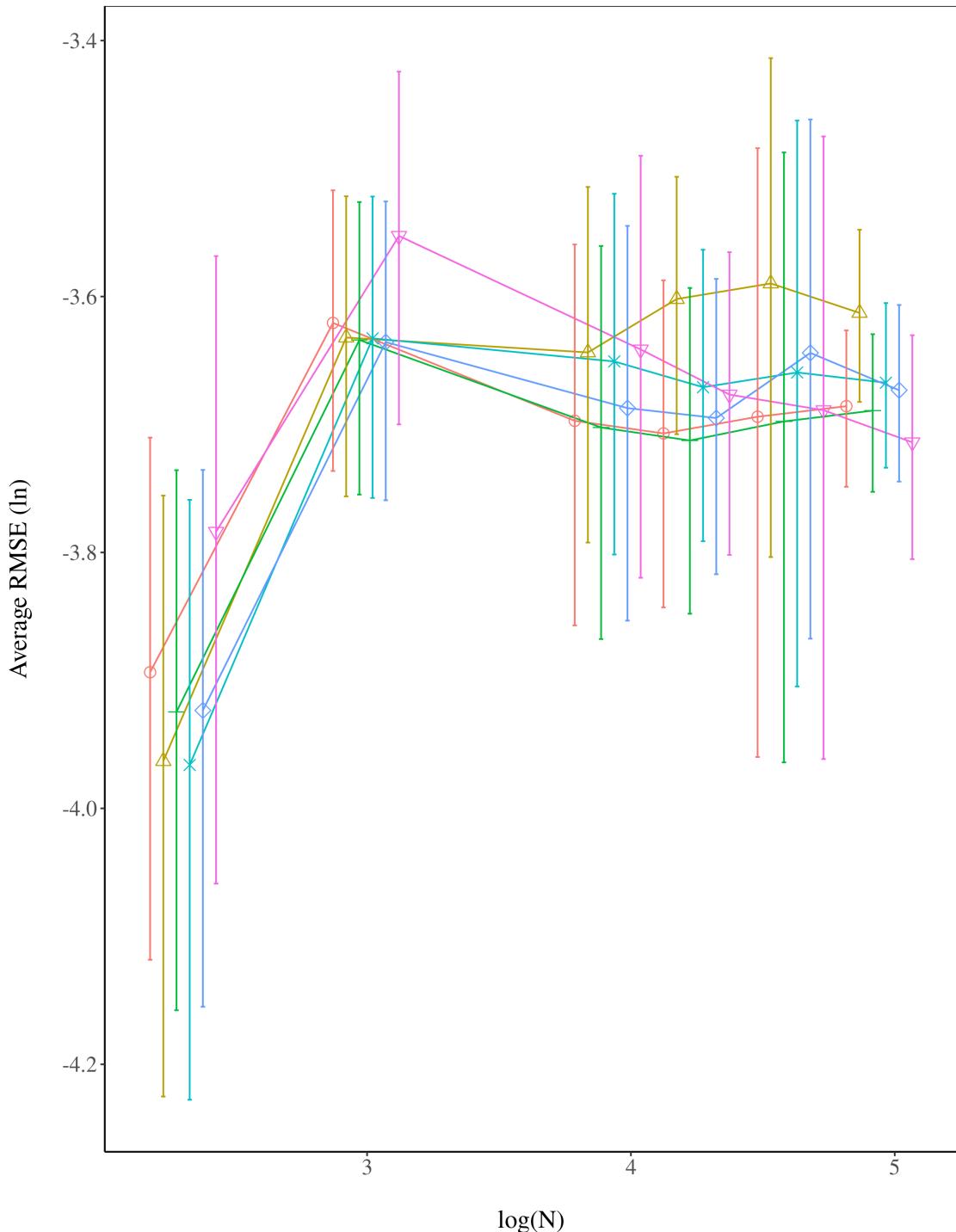


Figure 4.2: Placebo tests on stock market data: \ominus , DID; \triangleleft , ED; $+$, MC-NNM; \times , RVAE; \diamond , SCM; \triangleright , VT-EN.

compulsory primary education means to homogenize the population during the ‘Age of Mass Migration’, when of tens of millions of foreign migrants arrived to the country between 1850 and 1914.

An alternative explanation for the rise of public schooling is the view of Engerman and Sokoloff (2005) that frontier state governments sought to increase public investments in order to attract eastern migrants following the passage of the Homestead Act (HSA) of 1862, which opened for settlement hundreds of millions of acres of frontier land. Any adult citizen could apply for a homestead grant of 160 acres of land, provided that they live and make improvements on the land for five years. According to the authors, the sparse population on the frontier meant that state and local governments competed with each other to attract migrants in order to lower local labor costs and to increase land values and tax revenues. Frontier governments offered migrants broad access to cheap land and property rights, unrestricted voting rights, and a more generous provision of schooling and other public goods.

The HSA may have also increased state schooling expenditures by reducing the degree of land inequality on the frontier. Policies that led to the decentralization of public land are expected to lower land inequality by fixing land grants to 160 acres, thereby encouraging farm sizes to approach their ideal scale. Political economy frameworks (e.g., Acemoglu and Robinson 2008; Besley and Persson 2009) emphasize that greater economic power of the ruling class reduces public investments. In the model of Galor, Moav, and Vollrath (2009), wealthy landowners block education reforms because public schooling favors industrial labor productivity and decreases the value in farm rents. Inequality in this context can be thought of as a proxy for the amount of *de facto* political influence elites have to block reforms.

In the empirical application below, I apply the RNN-based approach to the problem of estimating the long-run impacts of the HSA on state government public education spending.

Data and assumptions

I create a state-level measure of state government education spending from the records of 48 state governments during the period of 1783 to 1932 (Sylla, Legler, and Wallis 1993) and the records of 16 state governments during the period of 1933 to 1937 (Sylla, Legler, and Wallis 1995a, 1995b). Comparable measures for 48 states are drawn from U.S. Census special reports for the years 1902, 1913, 1932, 1942, 1962, 1972, and 1982 (Haines 2010).

The data pre-processing steps are as follows. The measure is inflation-adjusted according to the U.S. Consumer Price Index (S. H. Williamson 2017) and scaled by the total free population in the decennial census (Haines 2010). Missing values are imputed separately in the pre- and -post-periods by carrying the last observation forward and remaining missing values are imputed by carrying the next observation backward. The data are log-transformed to alleviate exponential effects. Lastly, I remove states with no variance in the pre-period outcomes, resulting in a complete matrix of size $(N \times T) = (32 \times 156)$.

In this application, public land states — i.e., states crafted from the public domain — serve as treated units (i.e., the test set). State land states, which include states of the

original 13 colonies, Maine, Tennessee, Texas, Vermont, and West Virginia, were not directly affected by homestead policies and therefore serve as control units (i.e., the training set). The RNN-based approach assumes the distribution of $\mathbf{X}^{\text{train}}$ and \mathbf{X}^{test} are similar.

I weight the training loss by propensity scores in order to minimize distributional discrepancy between the training and test set inputs. The propensity scores are estimated via logistic regression with unit-specific, pre-period covariates including state-level average farm sizes measured in the 1860 and average farm values measured in the 1850 and 1860 censuses (Haines 2010) to control for homesteaders migrating to more productive land. To control for selection bias arising from differences in access to frontier lands, I create a measure of total miles of operational track per square mile aggregated to the state-level using digitized railroad maps provided by Atack (2013). Fig. SM-D.3 shows that the training and test set input distributions weighted by the propensity scores are visually similar.

Aggregating to the state level approximately 1.46 million individual land patent records authorized under the HSA, I determine that the earliest homestead entries occurred in 1869 in about half of the frontier states, about seven years following the enactment of the 1862 Homestead Act. Land patent records provide information on the initial transfer of land titles from the federal government and are made accessible online by the U.S. General Land Office (<https://glorecords.blm.gov>). Using this information, I set $T_0 = 87$, which leaves $T - T_0 = 69$ time periods when half of the states are exposed to treatment. While the approach assumes that treatment adoption is simultaneous across states, the date of initial treatment exposure varied as new frontier land opened between the period of 1869 to 1902. Also note that while the no interference assumption cannot directly be tested, it is likely that state land states were indirectly affected by the out-migration of homesteaders from frontier states.

Estimates

Prior to analyzing the data, I conduct placebo tests on the education spending data similar to those described in Section 4.4. Figure SM-D.4 presents the average RMSE calculated on the control unit outcomes with standard errors originating from 10 runs. In line with the previous placebo tests, the RNN-based estimators yield error rates comparable to the alternative estimators only when there are sufficient pre-period observations to train on; in this case, when $T_0/T \geq 0.5$. We can be reasonably confident that the RNN-based estimators will be at least as accurate as the other estimators since $T_0/T = 0.55$ in this application.

Next, I train a encoder-decoder network on the training set of state land states and use the learned weights to predict the counterfactual outcomes of public land states. The top panel of Figure 4.3 compares the average outcomes of treated units and control units along with the average predicted outcomes of treated units. The dashed vertical line represents the first year of treatment exposure in 1869. We are primarily interested in the difference in the observed and predicted treated unit outcomes, which is the quantity $\bar{\Phi}^{(t)}$. These per-period average causal impacts are plotted in the bottom panel and are bounded by 95%

randomization confidence intervals, which are estimated following the procedure described in Section SM-F.

Counterfactual predictions of state government education spending in the absence of the HSA generally tracks the observed control time-series until the turn of the 19th century, at which the counterfactual flattens and diverges from the increasing observed control time-series. This delay can potentially be explained by the fact that homestead entries did not substantially accumulate until after Congress prohibited the sale of public land in 1889 in all states except Missouri (Gates 1941, 1979).

Taking the mean of post-period impacts, I estimate that the impact of the HSA on the state government spending of states exposed to homesteads is 0.69 [-0.19, 2.01]. The confidence intervals surrounding this estimate contains zero, which implies that the estimated impact is not significantly more extreme than the exact distribution of average placebo effects under the null hypothesis. Examining the time-specific causal estimates reveals that fifty years after the first homestead entry, the estimated impact of the HSA on state government education spending in 1919 is 0.68 log points [0.13, 1.24]. The confidence intervals surrounding this time-specific estimate does not contain zero, which implies that the estimated impact is significantly more extreme than the average placebo effects. To put the magnitude of the point estimate in perspective, it represents about 3% of the total school expenditures per-capita in 1929 (Snyder and Dillow 2010).

4.6 Conclusion

This paper makes a methodological contribution in proposing a novel alternative to the SCM for estimating the effect of a policy intervention on an outcome over time. The SCM is growing in popularity in the social sciences despite its limitations — the most obvious being that the choice of specification can lead to different results, and thus facilitate *p*-hacking. By inputting only control unit outcomes and not relying on pre-period covariates, the proposed method offers a more principled approach than the SCM.

The RNN-based approach joins a new generation of data-driven machine learning techniques for generating counterfactual predictions. Machine learning techniques in general have an advantage over the SCM in that they automatically choose appropriate predictors without relying on pretreatment covariates; this capability limits “researcher degrees of freedom” that arises from choices on how to specify the model. RNNs do not assume a specific functional distribution, can learn nonconvex combinations of control units, and are specifically structured to exploit temporal dependencies in the data. RNNs are also capable of handling multiple treated units, which is useful because the model can share parameters across treated units, and thus generate more precise predictions in settings in which treated units share similar data-generating processes.

In placebo tests, RNN-based estimators perform comparatively worse than the alternatives on small-dimensional datasets such as those featured in the original synthetic control papers. Both RNN-based estimators require sufficient pre-period observations in order to

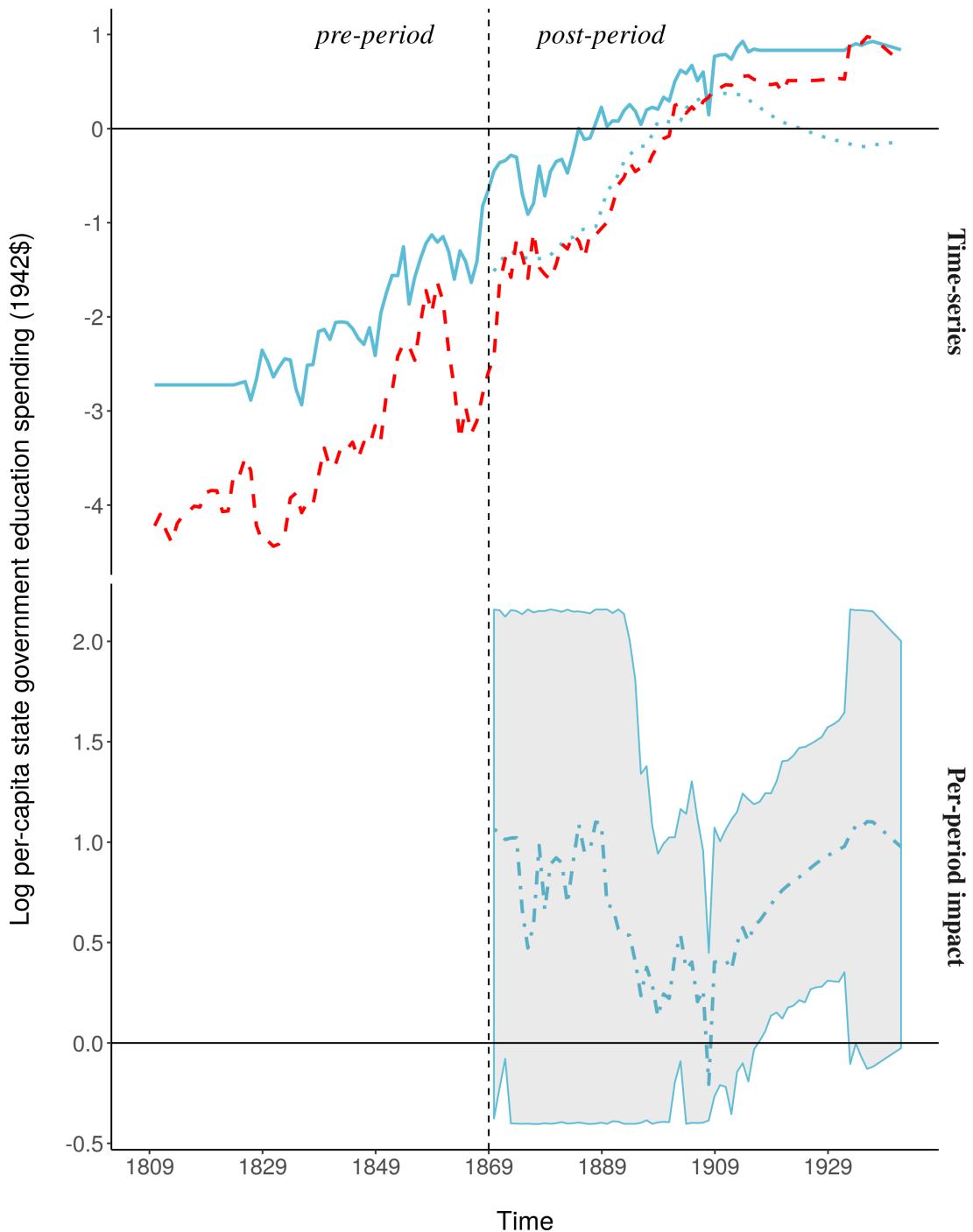


Figure 4.3: Encoder-decoder estimates of the impact of the HSA on state government education spending, 1809 to 1982: —, observed treated; -·-, observed control; ···, counterfactual treated; ——, $\bar{\Phi}^{(t)}$.

learn an informative representation of the control units. The RVAE in particular requires a large amount of training data since it is a self-supervised method that learns without outputs. In higher dimensional datasets such as the stock market data, the RNN-based methods generally outperform the alternatives when $N \ll T$. The estimators underperform when $N \gg T$, which again reflects the need for sufficient pre-period observations.

The matrix completion method performs well in either case, despite of its disadvantage of treating the data as static and thus ignoring the temporal component of the data. A built-in advantage of the matrix completion approach is that it does not assume a specific structure to the treatment assignment mechanism and thus can accommodate settings in which the time of initial treatment exposure varies across treated units. One potential avenue for future research is to integrate RNNs into the matrix completion approach by training multidirectional RNNs (e.g., Yoon, Zame, and Schaar 2018) to both impute missing values across the unit dimension and interpolate missing values within the time dimension.

A second area of future research would explore ways to relax the assumption of equivalence between the distributions of pre-period outcomes between control and treated units. An alternative approach to the one currently proposed is to treat the problem of counterfactual prediction like a NMT problem by training the networks on the pre-period outcomes of control units to predict those of treated units. The learned model weights would then be fit on the post-period outcomes of control units at test time. This setup would instead assume equivalence between the distributions of pre-and post-period outcomes of control units, which is more likely to be satisfied in the absence of interference between treated and control units.

Chapter 5

Conclusion

My dissertation argues that early land reforms had consequential long-run impacts on the development of the American state. The state-building role of land reform is frequently discussed in the context of comparative political economy, but rarely in the context of American political economy. I provide evidence that mid-nineteenth century homestead acts had positive long-run impacts that can help explain contemporary differences in the capacity of state governments.

My proposed research is to investigate the long-run environmental externalities of land reforms on the American frontier. Specifically, I will aggregate millions of individual settler decisions to the share of settled land and estimate the consequences of collective settler decisions on drought, soil erosion, and farm failure. This research will be informative for research in comparative political economy on the long-run environmental impacts of public land reform.

References

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. (2010). “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association* 105, no. 490 (2010): 493–505.
- . (2015). “Comparative Politics and the Synthetic Control Method.” *American Journal of Political Science* 59, no. 2 (2015): 495–510.
- Abadie, Alberto, and Javier Gardeazabal. (2003). “The Economic Costs of Conflict: A Case Study of the Basque Country.” *The American Economic Review* 93, no. 1 (2003): 113–132.
- Acemoglu, Daron, and James A Robinson. (2008). “Persistence of Power, Elites, and Institutions.” *American Economic Review* 98, no. 1 (2008): 267–293.
- Acemoglu, Daron, Davide Ticchi, and Andrea Vindigni. (2011). “Emergence and Persistence of Inefficient States.” *Journal of the European Economic Association* 9, no. 2 (2011): 177–208.
- Albertus, Michael. (2015). *Autocracy and Redistribution*. Cambridge: Cambridge University Press, 2015.
- Alesina, Alberto, and Bryony Reich. (2013). *Nation building*. Technical report. National Bureau of Economic Research, 2013.
- Allen, Douglas W. (1991). “Homesteading and Property Rights; Or, “How the West Was Really Won”.” *The Journal of Law and Economics* 34, no. 1 (1991): 1–23.
- Amjad, Muhammad, Devavrat Shah, and Dennis Shen. (2018). “Robust synthetic control.” *The Journal of Machine Learning Research* 19, no. 1 (2018): 802–852.
- Ansell, Ben, and David J Samuels. (2015). *Inequality and Democratization: An Elite Competition Approach*. Cambridge: Cambridge University Press, 2015.
- Archives, Georgia. (1978). *Georgia Official and Statistical Register 1977-1978*. Morrow, Georgia: Georgia Archives, 1978.

- Archives, Georgia. (1990). *Georgia Official and Statistical Register 1989-1990*. Morrow, Georgia: Georgia Archives, 1990.
- Atack, Jeremy. (2013). “On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited.” *The Journal of Economic History* 73, no. 2 (2013): 313–338.
- Atack, Jeremy, and Robert A Margo. (2011). “The Impact of Access to Rail Transportation on Agricultural Improvement: The American Midwest as a Test Case, 1850-1860.” *Journal of Transport and Land Use* 4, no. 2 (2011).
- Atack, Jeremy, Robert Margo, and Elisabeth Perlman. (2012). *The Impact of Railroads on School Enrollment in Nineteenth Century America*. Available from http://elisabethperlman.net/papers/PerlmanMargoAtack_SchoolRR.html, 2012.
- Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. (2017). “Matrix Completion Methods for Causal Panel Data Models.” *arXiv:1710.10251* (2017).
- Athey, Susan, and Guido Imbens. (2018). “Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption.” *arXiv:1808.05293* (2018).
- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. (2014). “Neural Machine Translation by Jointly Learning to Align and Translate.” *arXiv:1409.0473* (2014).
- Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo. (2018). “Nation-building through compulsory schooling during the age of mass migration.” *The Economic Journal* 129, no. 617 (2018): 62–109.
- Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse. (2017). *Frontier Culture: The Roots and Persistence of “Rugged Individualism” in the United States*. Technical report. National Bureau of Economic Research, 2017.
- Ben-David, Shai, John Blitzer, Koby Crammer, and Fernando Pereira. (2007). “Analysis of representations for domain adaptation.” In *Advances in Neural Information Processing Systems*, 137–144. 2007.
- Benabou, Roland. (2000). “Unequal Societies: Income Distribution and the Social Contract.” *American Economic Review* (2000): 96–129.
- Bensel, Richard Franklin. (1990). *Yankee Leviathan: the Origins of Central State Authority in America, 1859-1877*. Cambridge: Cambridge University Press, 1990.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. (2004). “How Much Should We Trust Differences-in-Differences Estimates?” *The Quarterly Journal of Economics* 119, no. 1 (2004): 249–275.

- Besley, Timothy, and Stephen Coate. (1997). “An Economic Model of Representative Democracy.” *The Quarterly Journal of Economics* 112, no. 1 (1997): 85–114.
- Besley, Timothy, and Torsten Persson. (2009). “The Origins of State Capacity: Property Rights, Taxation and Politics.” *American Economic Review* 99, no. 4 (2009): 1218–1244.
- . (2010). “State Capacity, Conflict, and Development.” *Econometrica* 78, no. 1 (2010): 1–34.
- Bickel, Steffen, Michael Brückner, and Tobias Scheffer. (2007). “Discriminative learning for differing training and test distributions.” In *Proceedings of the 24th international conference on Machine learning*, 81–88. ACM, 2007.
- . (2009). “Discriminative learning under covariate shift.” *Journal of Machine Learning Research* 10, no. Sep (2009): 2137–2155.
- Bleakley, Hoyt, and Joseph Ferrie. (2016). “Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital Across Generations.” *The Quarterly Journal of Economics* (2016): qjw014.
- Bleakley, Hoyt, and Joseph P Ferrie. (2013). “Up from Poverty? The 1832 Cherokee Land Lottery and the Long-run Distribution of Wealth.” 2013.
- Boix, Carles. (2003). *Democracy and Redistribution*. Cambridge: Cambridge University Press, 2003.
- Bourgignon, François, and Thierry Verdier. (2000). “Oligarchy, Democracy, Inequality and Growth.” *Journal of Development Economics* 62, no. 2 (2000): 285–313.
- Bowman, Samuel R, Luke Vilnis, Oriol Vinyals, Andrew M Dai, Rafal Jozefowicz, and Samy Bengio. (2015). “Generating Sentences from a Continuous Space.” *arXiv:1511.06349* (2015).
- Cadle, Farris W. (1991). *Georgia Land Surveying History and Law*. University of Georgia Press, 1991.
- Candes, Emmanuel J, and Yaniv Plan. (2010). “Matrix completion with noise.” *Proceedings of the IEEE* 98, no. 6 (2010): 925–936.
- Candès, Emmanuel J, and Benjamin Recht. (2009). “Exact matrix completion via convex optimization.” *Foundations of Computational mathematics* 9, no. 6 (2009): 717.
- Carpenter, Daniel P. (2001). *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton, NJ: Princeton University Press, 2001.

- Carvalho, Carlos, Ricardo Masini, and Marcelo C Medeiros. (2018). “ArCo: An artificial counterfactual approach for high-dimensional panel time-series data.” *Journal of Econometrics* 207, no. 2 (2018): 352–380.
- Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. (2013). “Catastrophic Natural Disasters and Economic Growth.” *Review of Economics and Statistics* 95, no. 5 (2013): 1549–1561.
- Chernozhukov, Victor, Kaspar Wuthrich, and Yinchu Zhu. (2017). “An exact and robust conformal inference method for counterfactual and synthetic controls.” *arXiv:1712.09089* (2017).
- Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. (2014). “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.” *arXiv:1406.1078* (2014).
- Chollet, François. (2015). *Keras*. Available at: <https://github.com/fchollet/keras>, 2015.
- Chorowski, Jan K, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. (2015). “Attention-Based Models for Speech Recognition.” In *Advances in Neural Information Processing Systems*, 577–585. 2015.
- Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. (2014). “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.” *arXiv:1412.3555* (2014).
- Chung, Junyoung, Kyle Kastner, Laurent Dinh, Kratarth Goel, Aaron C Courville, and Yoshua Bengio. (2015). “A Recurrent Latent Variable Model for Sequential Data.” In *Advances in Neural Information Processing Systems*, 2980–2988. 2015.
- Cinar, Yagmur Gizem, Hamid Mirsaee, Parantapa Goswami, Eric Gaussier, Ali Aït-Bachir, and Vadim Strijov. (2017). “Position-Based Content Attention for Time Series Forecasting with Sequence-to-Sequence RNNs.” In *International Conference on Neural Information Processing*, 533–544. Springer, 2017.
- Clayton, Augustin Smith, and Thomas W Adams. (1812). *A Compilation of the Laws of the State of Georgia: Passed by the Legislature Since the Political Year 1800, to the Year 1810, Inclusive*. Adams & Duyckinck, 1812.
- Cortes, Corinna, Yishay Mansour, and Mehryar Mohri. (2010). “Learning bounds for importance weighting.” In *Advances in Neural Information Processing Systems*, 442–450. 2010.
- Cortes, Corinna, Mehryar Mohri, Michael Riley, and Afshin Rostamizadeh. (2008). “Sample selection bias correction theory.” In *International Conference on Algorithmic Learning Theory*, 38–53. Springer, 2008.

- Corvalan, Alejandro, Pablo Querubin, and Sergio Vicente. (forthcoming, 2018). "The Political Class and Redistributive Policies." *Journal of the European Economic Association* (forthcoming, 2018).
- Coulter, Ellis Merton. (1960). *Georgia: A Short History*. University of North Carolina Press, 1960.
- Davis, Robert Scott. (1981). *Research in Georgia*. Southern Historical Press, 1981.
- Doherty, Daniel, Alan S Gerber, and Donald P Green. (2006). "Personal Income and Attitudes Toward Redistribution: A Study of Lottery Winners." *Political Psychology* 27, no. 3 (2006): 441–458.
- Donaldson, Dave, and Richard Hornbeck. (2016). "Railroads and American Economic Growth: A "Market Access" Approach." *The Quarterly Journal of Economics* 131, no. 2 (2016): 799–858.
- Doudchenko, Nikolay, and Guido W Imbens. (2016). "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis." *arXiv:1610.07748* (2016).
- Dube, Arindrajit, and Ben Zipperer. (2015). *Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies*. IZA Discussion Paper No. 8944. Available at: <http://ftp.iza.org/dp8944.pdf>, 2015.
- Eggers, Andrew C, and Jens Hainmueller. (2009). "MPs for sale? Returns to office in postwar British politics." *American Political Science Review* 103, no. 04 (2009): 513–533.
- El Hihi, Salah, and Yoshua Bengio. (1995). "Hierarchical Recurrent Neural Networks for Long-Term Dependencies." In *Neural Information Processing Systems*, 400:409. 1995.
- Elster, Jon. (1978). *Logic and Society Contradictions and Possible Worlds*. John Wiley & Sons, 1978.
- Engerman, Stanley L, and Kenneth L Sokoloff. (2005). "The Evolution of Suffrage Institutions in the New World." *The Journal of Economic History* 65, no. 4 (2005): 891–921.
- Fabius, Otto, and Joost R van Amersfoort. (2014). "Variational recurrent auto-encoders." *arXiv:1412.6581* (2014).
- Fearon, James D. (1991). "Counterfactuals and hypothesis testing in political science." *World Politics* 43, no. 2 (1991): 169–195.
- Ferman, Bruno, and Cristine Pinto. (2016). *Revisiting the Synthetic Control Estimator*. Available at <https://mpra.ub.uni-muenchen.de/81941/>, 2016.
- _____. (2018). *Synthetic Controls with Imperfect Pre-Treatment Fit*. Available at: <https://sites.google.com/site/brunoferman/research>, 2018.

- Ferman, Bruno, Cristine Pinto, and Vitor Possebom. (2018). *Cherry picking with synthetic controls*. Available at: https://mpra.ub.uni-muenchen.de/85138/1/MPRA_paper_85138.pdf, 2018.
- Ferrie, Joseph P. (1997). “Migration to the Frontier in Mid-Nineteenth Century America: A Re-Examination of Turner’s ‘Safety Valve’.” 1997.
- Fiorina, Morris P. (1994). “Divided Government in the American States: A Byproduct of Legislative Professionalism?” *American Political Science Review* (1994): 304–316.
- . (1999). “Further Evidence of the Partisan Consequences of Legislative Professionalism.” *American Journal of Political Science* (1999): 974–977.
- Firpo, Sergio, and Vitor Possebom. (2018). “Synthetic control method: Inference, sensitivity analysis and confidence sets.” *Journal of Causal Inference* 6, no. 2 (2018).
- Fisher, Ronald Aylmer. (1935). *The Design of Experiments*. Oliver & Boyd, 1935.
- Fogel, Robert William. (1964). *Railroads and American Economic Growth*. Johns Hopkins Press: Baltimore, 1964.
- Frymer, Paul. (2014). “‘A Rush and a Push and the Land Is Ours’: Territorial Expansion, Land Policy, and U.S. State Formation.” *Perspectives on Politics* 12, no. 1 (2014): 119.
- Galor, Oded, Omer Moav, and Dietrich Vollrath. (2009). “Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence.” *The Review of Economic Studies* 76, no. 1 (2009): 143–179.
- Ganin, Yaroslav, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. (2015). “Domain-Adversarial Training of Neural Networks.” *arXiv e-prints*, arXiv:1505.07818 (May 2015): arXiv:1505.07818. arXiv: 1505.07818 [stat.ML].
- García-Jimeno, Camilo, and James A Robinson. (2008). “The myth of the frontier.” In *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*, 49–88. Chicago, IL: University of Chicago Press, 2008.
- Gates, Paul W. (1939). “Land Policy and Tenancy in the Prairie Counties of Indiana.” *The Indiana Magazine of History* (1939): 1–26.
- . (1940). “Federal Land Policy in the South 1866-1888.” *The Journal of Southern History* 6, no. 3 (1940): 303–330.
- . (1941). “Land Policy and Tenancy in the Prairie States.” *The Journal of Economic History* 1, no. 1 (1941): 60–82.
- . (1942). “The Role of the Land Speculator in Western Development.” *The Pennsylvania Magazine of History and Biography* 66, no. 3 (1942): 314–333.

- Gates, Paul W. (1979). "Federal Land Policies in the Southern Public Land States." *Agricultural History* 53, no. 1 (1979): 206–227.
- Glorot, Xavier, and Yoshua Bengio. (2010). "Understanding the Difficulty of Training Deep Feedforward Neural Networks." In *Artificial Intelligence and Statistics*, 9:249–256. 2010.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. (2016). *Deep Learning*. Cambridge, MA: MIT press, 2016.
- Graham, Paul K. (2004). *1805 Georgia Land Lottery Fortunate Drawers and Grantees*. Genealogy Company, 2004.
- . (2011). *1807 Georgia Land Lottery Fortunate Drawers and Grantees*. Genealogy Company, 2011.
- Graves, Alex. (2012). "Neural Networks." In *Supervised Sequence Labelling with Recurrent Neural Networks*, 15–35. Springer, 2012.
- Groseclose, Timothy, and Keith Krehbiel. (1994). "Golden Parachutes, Rubber Checks, and Strategic Retirements from the 102d House." *American Journal of Political Science* 38, no. 1 (1994): 75–99.
- Hacker, J David. (2010). "Decennial Life Tables for the White Population of the United States, 1790–1900." *Historical Methods* 43, no. 2 (2010): 45–79.
- Hahn, Jinyong, and Ruoyao Shi. (2017). "Synthetic control and inference." *Econometrics* 5, no. 4 (2017): 52.
- Haines, Michael R. (2004). *The Inter-University Consortium for Political and Social Research. Historical, Demographic, Economic, and Social Data: The United States, 1790-2000 [Computer file]*. ICPSR02896-v2. Hamilton, NY: Colgate University. Ann Arbor: MI: Inter-University Consortium for Political and Social Research, 2004.
- . (2010). *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. doi.org/10.3886/ICPSR02896.v3, 2010.
- Hall, Richard L, and Robert P Van Houweling. (1995). "Avarice and Ambition in Congress: Representatives' Decisions to Run or Retire from the U.S. House." *American Political Science Review* 89, no. 01 (1995): 121–136.
- Heckman, James J. (1979). "Sample selection bias as a specification error." *Econometrica: Journal of the Econometric Society* (1979): 153–161.
- Helper, Hinton Rowan. (1860). *The Impending Crisis of the South: How to Meet It*. AB Burdick, 1860.

- Huang, Jiayuan, Arthur Gretton, Karsten M Borgwardt, Bernhard Schölkopf, and Alex J Smola. (2007). "Correcting sample selection bias by unlabeled data." In *Advances in Neural Information Processing Systems*, 601–608. 2007.
- ICPSR. (1984). *Candidate Name and Constituency Totals, 1788-1984*. <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/2>. [Online; accessed 06-June-2016], 1984.
- Ilin, Alexander, and Tapani Raiko. (2010). "Practical approaches to principal component analysis in the presence of missing values." *Journal of Machine Learning Research* 11, no. Jul (2010): 1957–2000.
- Imbens, Guido W, and Donald B Rubin. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press, 2015.
- Johansson, Fredrik, Uri Shalit, and David Sontag. (2016). "Learning representations for counterfactual inference." In *International Conference on Machine Learning*, 3020–3029. 2016.
- Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler. (2015). *Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates*. Available at: http://www.oekonometrie.uni-saarland.de/papers/SCM_Predictors.pdf, 2015.
- Kingma, Diederik P, and Max Welling. (2013). "Auto-encoding variational bayes." *arXiv:1312.6114* (2013).
- Kingma, Diederik, and Jimmy Ba. (2014). "Adam: A Method for Stochastic Optimization." *arXiv:1412.6980* (2014).
- Klößner, Stefan, Ashok Kaul, Gregor Pfeifer, and Manuel Schieler. (2017). "Comparative Politics and the Synthetic Control Method Revisited: A Note on Abadie et al. (2015)." *Swiss Journal of Economics and Statistics* (2017).
- Kotlikoff, Laurence J. (1979). "The Structure of Slave Prices in New Orleans, 1804 to 1862." *Economic Inquiry* 17, no. 4 (1979): 496–518.
- Lamb, Alex M, Anirudh Goyal Alias Parth Goyal, Ying Zhang, Saizheng Zhang, Aaron C Courville, and Yoshua Bengio. (2016). "Professor forcing: A new algorithm for training recurrent networks." In *Advances In Neural Information Processing Systems*, 4601–4609. 2016.
- Lampi, Philip. (2013). *A New Nation Votes: American Election Returns, 1787-1825*. <http://elections.lib.tufts.edu/>. [Online; accessed 06-June-2016], 2013.
- Lanza, Michael L. (1999). *Agrarianism and Reconstruction Politics: The Southern Homestead Act*. Baton Rouge, LA: LSU Press, 1999.
- Lewis, David. (2013). *Counterfactuals*. John Wiley & Sons, 2013.

- Little, Roderick JA, and Donald B Rubin. (2014). *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley & Sons, 2014.
- Mares, Isabela, and Didac Queralt. (2015). “The Non-Democratic Origins of Income Taxation.” *Comparative Political Studies* 48, no. 14 (2015): 1974–2009.
- Mazumder, Rahul, Trevor Hastie, and Robert Tibshirani. (2010). “Spectral regularization algorithms for learning large incomplete matrices.” *Journal of Machine Learning Research* 11, no. Aug (2010): 2287–2322.
- Meltzer, Allan H, and Scott F Richard. (1981). “A Rational Theory of the Size of Government.” *Journal of Political Economy* 89, no. 5 (1981): 914–927.
- Meyer, John W, David Tyack, Joane Nagel, and Audri Gordon. (1979). “Public education as nation-building in America: Enrollments and bureaucratization in the American states, 1870-1930.” *American Journal of Sociology* 85, no. 3 (1979): 591–613.
- Meyers, Christopher C, and David Williams. (2012). *Georgia: A Brief History*. Mercer University Press, 2012.
- Milyo, Jeffrey, and Timothy Groseclose. (1999). “The Electoral Effects of Incumbent Wealth.” *The Journal of Law and Economics* 42, no. 2 (1999): 699–722.
- Murtazashvili, Ilia. (2013). *The Political Economy of the American Frontier*. Cambridge: Cambridge University Press, 2013.
- Murtazashvili, Ilia, and Jennifer Murtazashvili. (2016). “Does the Sequence of Land Reform and Political Reform Matter? Evidence from State-Building in Afghanistan.” *Conflict, Security & Development* 16, no. 2 (2016): 145–172.
- Neyman, Jerzy. (1923). “On the application of probability theory to agricultural experiments.” Reprinted in Splawa-Neyman, Dabrowska, Speed, et al. 1990, *Annals of Agricultural Sciences* 51, no. 1 (1923).
- Novak, William J. (2008). “The Myth of the “Weak” American State.” *The American Historical Review* 113, no. 3 (2008): 752–772.
- Officer, Lawrence H, and S Williamson. (2012). *The Annual Consumer Price Index for the United States, 1774-2010*. <http://www.whatisitworth.com>, 2012.
- Osborne, Martin J, and Al Slivinski. (1996). “A model of political competition with citizen-candidates.” *The Quarterly Journal of Economics* 111, no. 1 (1996): 65–96.
- Pearl, Judea. (2009). *Causality*. Cambridge University Press, 2009.
- Perotti, Roberto. (1993). “Political Equilibrium, Income Distribution, and Growth.” *The Review of Economic Studies* 60, no. 4 (1993): 755–776.

- Phillips, Ulrich B. (1905). "The economic cost of slaveholding in the cotton belt." *Political Science Quarterly* 20, no. 2 (1905): 257–275.
- Politis, Dimitris N, and Halbert White. (2004). "Automatic Block-Length Selection for the Dependent Bootstrap." *Econometric Reviews* 23, no. 1 (2004): 53–70.
- Poulos, Jason. (2017). "RNN-based counterfactual time-series prediction." *arXiv:1712.03553* (2017).
- . (2019). "Land lotteries, long-term wealth, and political selection." *Public Choice* 178, no. 1 (2019): 217–230.
- Querubin, Pablo, James M Snyder Jr, et al. (2013). "The Control of Politicians in Normal Times and Times of Crisis: Wealth Accumulation by US Congressmen, 1850–1880." *Quarterly Journal of Political Science* 8, no. 4 (2013): 409–450.
- Ramcharan, Rodney. (2010). "Inequality and Redistribution: Evidence from U.S. Counties and States, 1890–1930." *The Review of Economics and Statistics* 92, no. 4 (2010): 729–744.
- Rosenbaum, Paul R, and Donald B Rubin. (1983). "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70, no. 1 (1983): 41–55.
- Rossi, Martín A. (2014). "The Impact of Individual Wealth on Posterior Political Power." *Journal of Economic Behavior & Organization* 106, no. C (2014): 469–480.
- Rubin, Donald B. (1990). "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science* 5, no. 4 (1990): 472–480.
- Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. (2015). *Integrated Public Use Microdata Series: Version 6.0 [database]*. <https://www.nappdata.org/>. [Online; accessed 06-June-2016], 2015.
- Saint-Paul, Gilles, and Thierry Verdier. (1993). "Education, Democracy and Growth." *Journal of Development Economics* 42, no. 2 (1993): 399–407.
- Schmidhuber, Jürgen, and Sepp Hochreiter. (1997). "Long Short-Term Memory." *Neural Computation* 9, no. 8 (1997): 1735–1780.
- Schnabel, Tobias, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. (2016). "Recommendations as treatments: Debiasing learning and evaluation." *arXiv:1602.05352* (2016).
- Shanks, Trina RW. (2005). "The Homestead Act: A Major Asset-Building Policy in American History." *Inclusion in the American Dream: Assets, Poverty, and Public Policy* (2005): 20–41.
- Simons, Algie Martin. (1912). *Social Forces in American History*. Macmillan, 1912.

- Skowronek, Stephen. (1982). *Building a New American State: The Expansion of National Administrative Capacities, 1877-1920*. Cambridge: Cambridge University Press, 1982.
- Snyder, Thomas D, and Sally A Dillow. (2010). “Digest of Education Statistics, 2009. NCES 2010-013.” *National Center for Education Statistics* (2010).
- Splawa-Neyman, Jerzy, DM Dabrowska, TP Speed, et al. (1990). “On the application of probability theory to agricultural experiments. Essay on principles. Section 9.” *Statistical Science* 5, no. 4 (1990): 465–472.
- Suryanarayan, Pavithra. (2017). “Hollowing Out the State: Franchise Expansion and Fiscal Capacity in Colonial India.” Available from <https://ssrn.com/abstract=2951947>, 2017.
- Sylla, Richard E, John B Legler, and John Wallis. (1993). *Sources and Uses of Funds in State and Local Governments, 1790-1915: [United States]*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2017-05-21. doi.org/10.3886/ICPSR06304.v1, 1993.
- _____. (1995a). *State and Local Government [United States]: Sources and Uses of Funds, Census Statistics, Twentieth Century [Through 1982]*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2017-05-21. doi.org/10.3886/ICPSR06304.v1, 1995.
- _____. (1995b). *State and Local Government [United States]: Sources and Uses of Funds, State Financial Statistics, 1933-1937*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2017-05-21. http://doi.org/10.3886/ICPSR06306.v1, 1995.
- Sylla, Richard, and John Joseph Wallis. (1998). “The Anatomy of Sovereign Debt Crises: Lessons from the American State Defaults of the 1840s.” *Japan and the World Economy* 10, no. 3 (1998): 267–293.
- Tetlock, Philip E, and Aaron Belkin. (1996). *Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives*. Princeton University Press, 1996.
- Troyanskaya, Olga, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. (2001). “Missing value estimation methods for DNA microarrays.” *Bioinformatics* 17, no. 6 (2001): 520–525.
- Truex, Rory. (2014). “The Returns to Office in a “Rubber Stamp” Parliament.” *American Political Science Review* 108, no. 02 (2014): 235–251.
- Turner, Frederick Jackson. (1956). *The Significance of the Frontier in American History*. Ithaca, NY: Cornell University Press, 1956.
- Vinyals, Oriol, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. (2014). “Grammar as a Foreign Language.” *arXiv:1412.7449* (2014).

- Vollrath, Dietrich. (2013). "Inequality and School Funding in the Rural United States, 1890." *Explorations in Economic History* 50, no. 2 (2013): 267–284.
- Wallis, John Joseph. (2000). "American Government Finance in the Long Run: 1790 to 1990." *The Journal of Economic Perspectives* 14, no. 1 (2000): 61–82.
- Wallis, John Joseph, Richard E Sylla, and Arthur Grinath III. (2004). *Sovereign Debt and Repudiation: The Emerging-Market Debt Crisis in the US States, 1839-1843*, 2004.
- Weiman, David F. (1991). "Peopling the Land by Lottery? The Market in Public Lands and the Regional Differentiation of Territory on the Georgia Frontier." *The Journal of Economic History* 51, no. 04 (1991): 835–860.
- Williamson, Jeffrey G. (2008). *Late Nineteenth-Century American Development: A General Equilibrium History*. Cambridge University Press, 2008.
- Williamson, Samuel H. (2017). *Seven ways to compute the relative value of a U.S. dollar amount, 1774 to present*. Available from <http://MeasuringWorth.com>, 2017.
- . (2018). *The Annual Consumer Price Index for the United States, 1774-2015*. <https://www.measuringworth.com/uscp/>, 2018.
- Winters, Jeffrey A, and Benjamin I Page. (2009). "Oligarchy in the United States?" *Perspectives on Politics* 7, no. 04 (2009): 731–751.
- Wright, Gavin. (1986). "Old South, New South." NY: Basic Books (1986).
- Xu, Yiqing. (2017). "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25, no. 1 (2017): 57–76.
- Yoon, Jinsung, William R Zame, and Mihaela van der Schaar. (2018). "Estimating missing data in temporal data streams using multi-directional recurrent neural networks." *IEEE Transactions on Biomedical Engineering* (2018).
- Zhu, Lingxue, and Nikolay Laptev. (2017). "Deep and Confident Prediction for Time Series at Uber." *arXiv:1709.01907* (2017).
- Ziblatt, Daniel. (2008). "Does Landholding Inequality Block Democratization? A Test of the "Bread and Democracy" Thesis and the Case of Prussia." *World Politics* 60, no. 4 (2008): 610–641.
- Zou, Hui, and Trevor Hastie. (2005). "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, no. 2 (2005): 301–320.

Appendix A

Supplementary Tables and Figures for Chapter 2

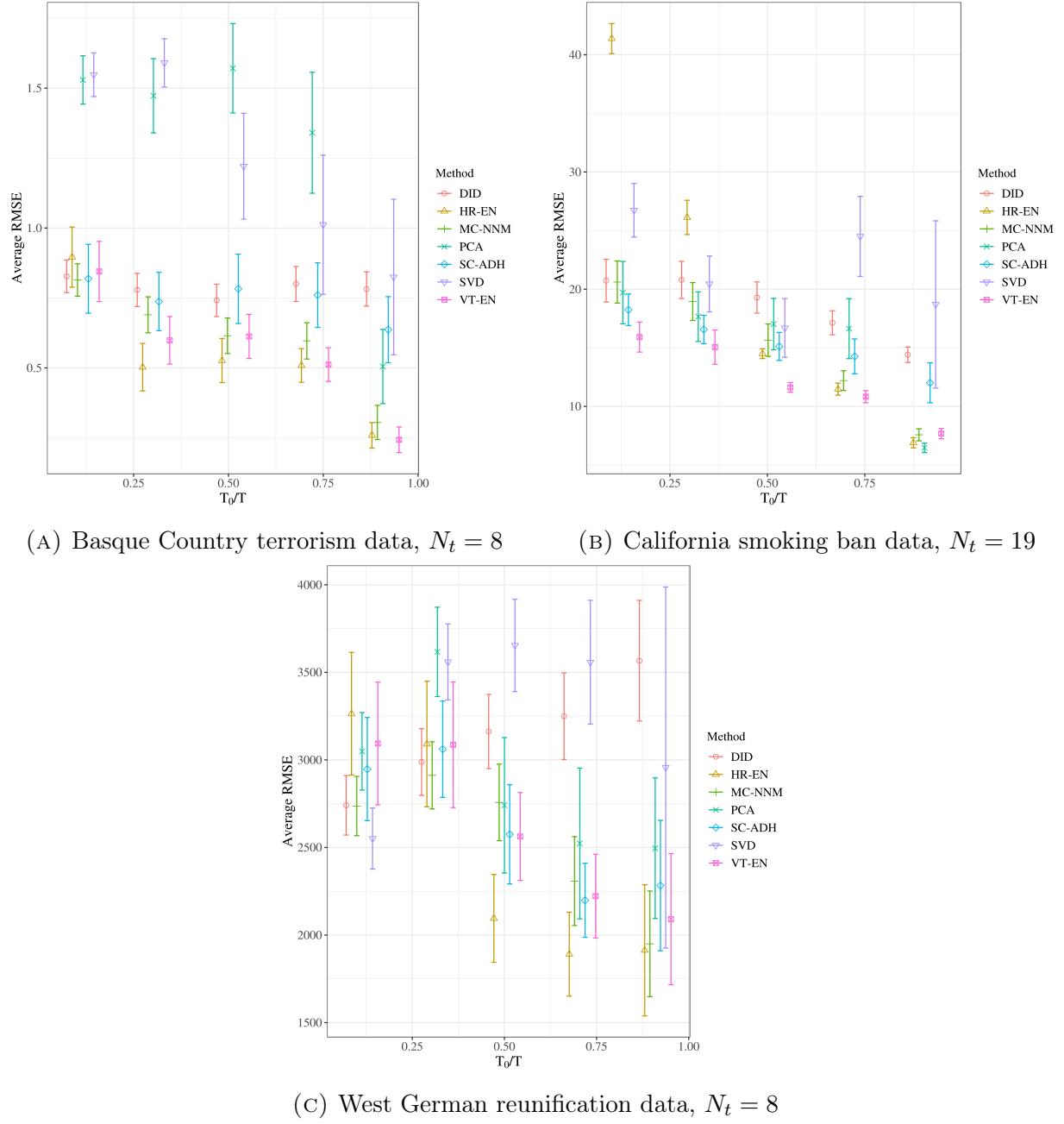


Figure A.1: Placebo tests under simultaneous treatment adoption.

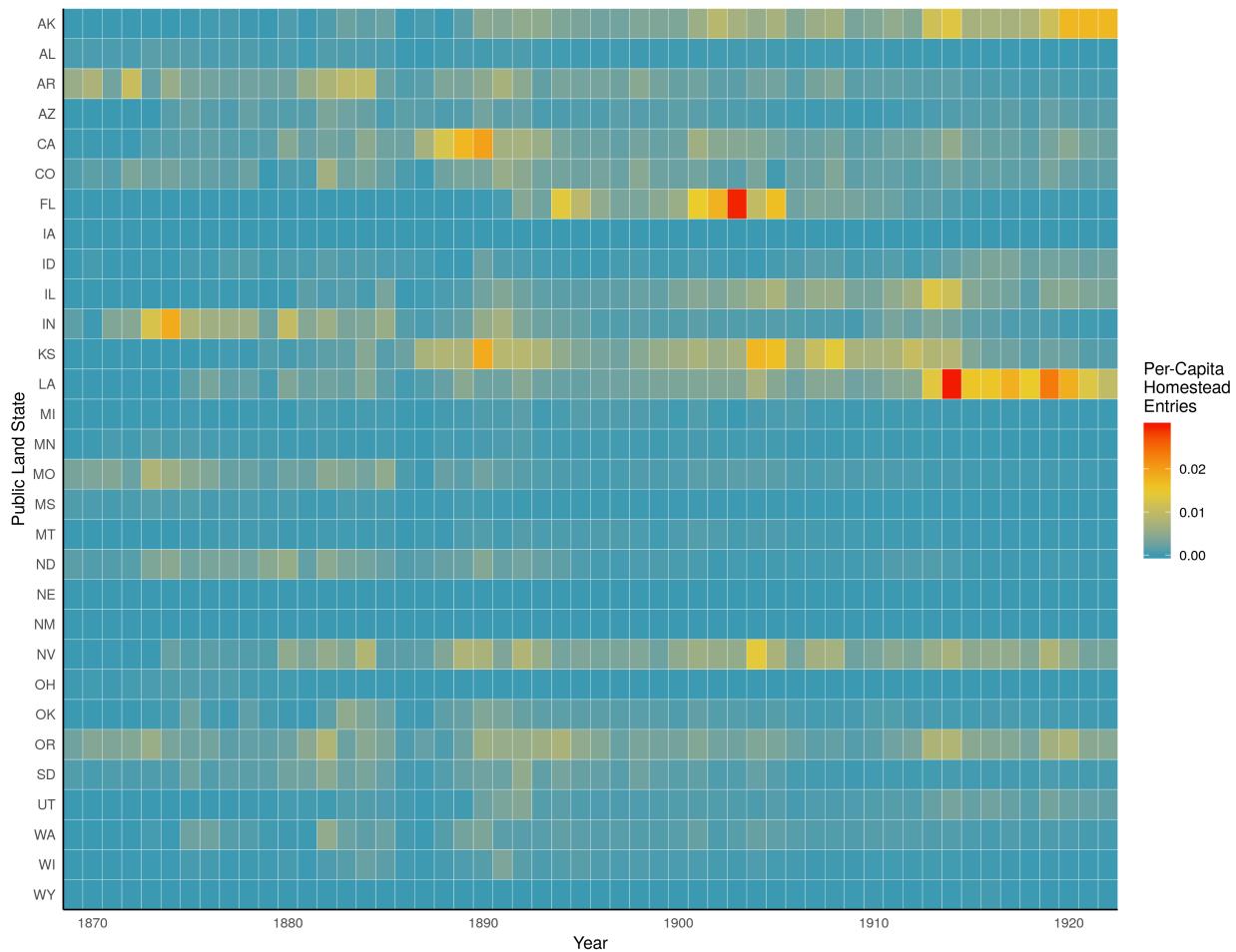


Figure A.2: Per-capita homestead entries in state i and year t , 1869-1922.

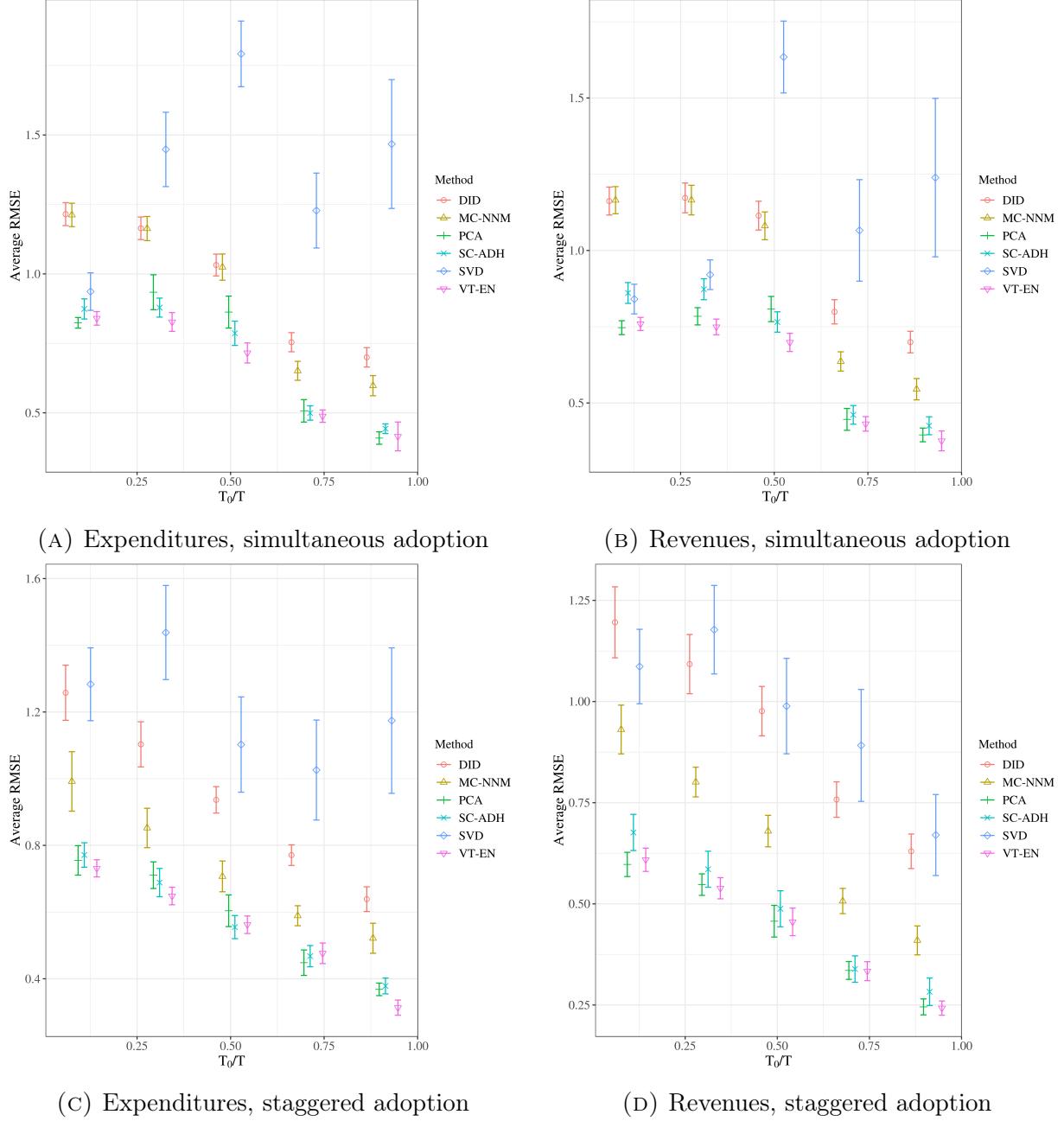


Figure A.3: Placebo tests under simultaneous and staggered treatment adoption, with $N_t = 9$.

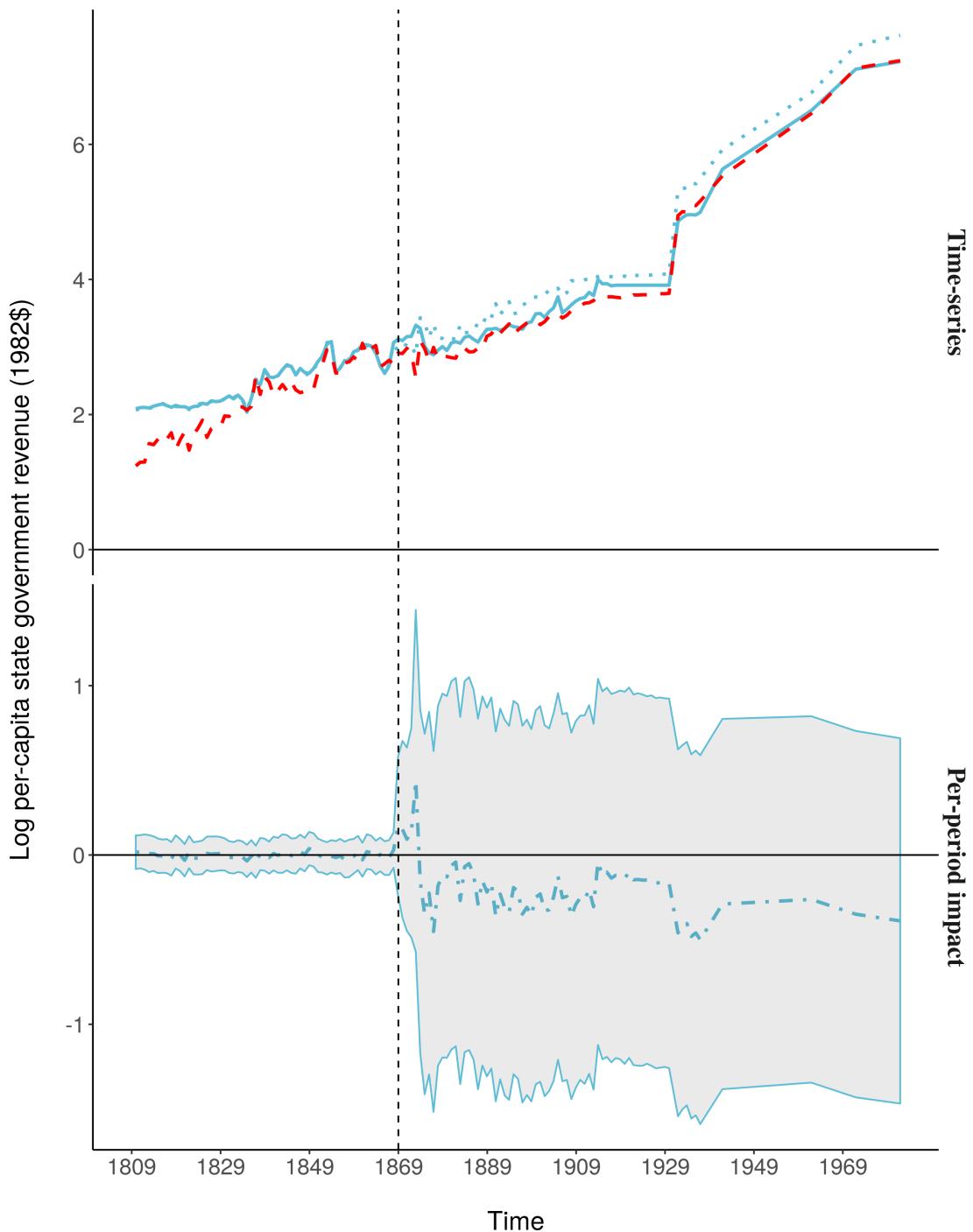


Figure A.4: MC-NNM estimates of treatment exposure on state government revenue, 1809 to 1982: —, observed treated; - - -, observed control; , counterfactual treated; - - - - , $\hat{\alpha}_t$.

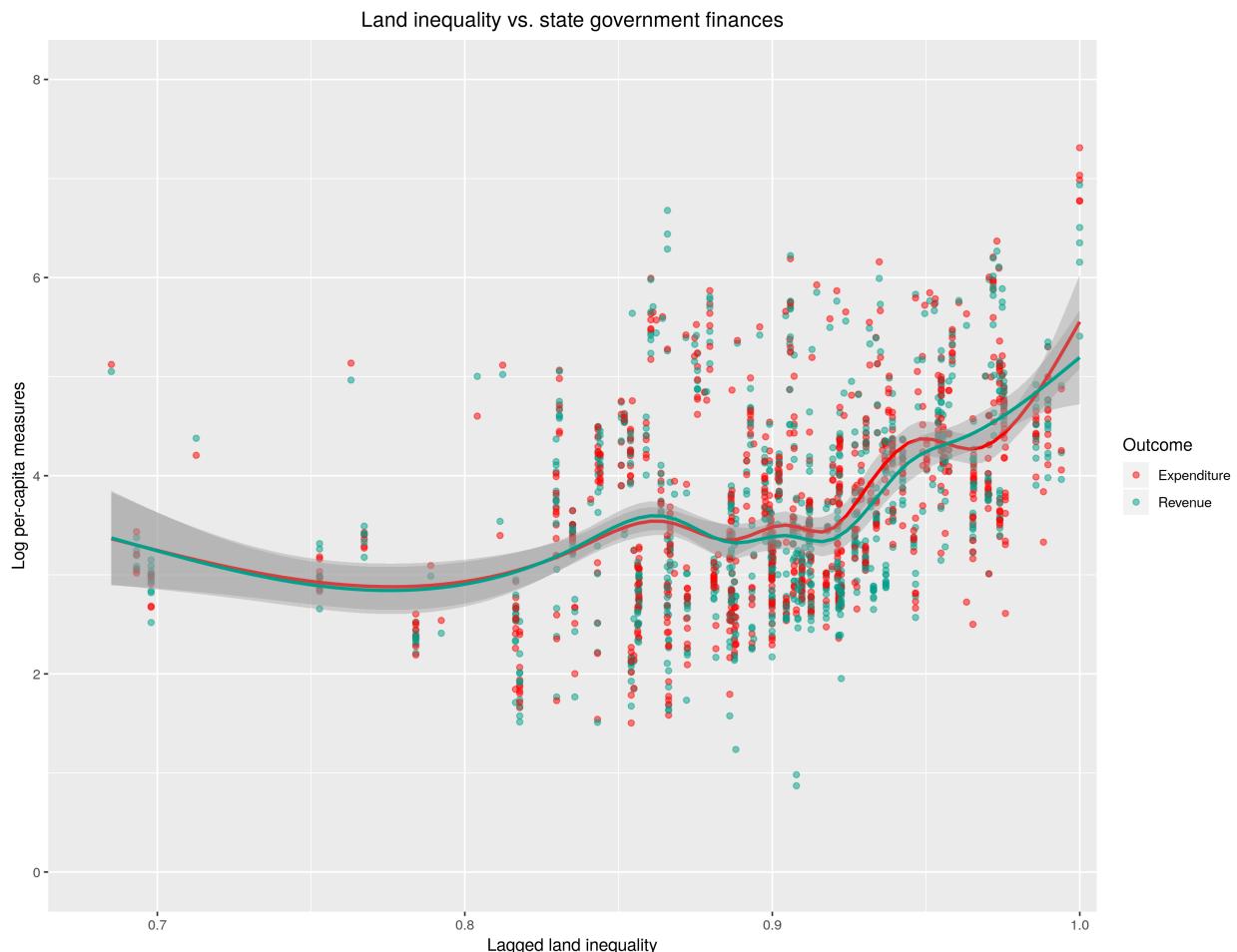


Figure A.5: Land inequality (lagged by 10 years) vs. log per-capita revenue and expenditure, 1860-1950. Each point is a state-year observation. Lines represent generalized additive model (GAM) fits to the data and shaded regions represent corresponding 95% confidence intervals.

Appendix B

Supplementary Tables and Figures for Chapter 3

Table B.1: Counties created by 1805 and 1807 lotteries.

Panel A: 1805						
Counties	No. Districts	Lot sizes (acres)	Lot length (chains square)	Lot orientation (degrees)	Grant fee (\$)	Est. value of lot (\$)
Baldwin	5	202.5	45	45 / 60	8.10	839.17
Wayne	3	490	70	13 / 77	19.60	842.64
Wilkinson	5	202.5	45	45 / 60	8.10	811.25

Panel B: 1807						
Counties	No. Districts	Lot sizes (acres)	Lot length (chains square)	Lot orientation (degrees)	Grant fee (\$)	Est. value of lot (\$)
Baldwin	15	202.5	45	45 / 60	12.15	827.35
Wilkinson	23	202.5	45	45 / 60	12.15	799.82

Notes: counties and land lots specified by Acts of 11 May 1803 and 9 June 1806. Lot orientation is degrees from the meridian. Lot values are estimated by averaging the cash value of farms minus the value of farming implements and machinery by the number of (improved and unimproved) acres of land in farms (Haines 2004; Bleakley and J. P. Ferrie 2013). The 1850 values are deflated to 1805 dollars (Panel A) and 1807 dollars (Panel B) using a historical consumer price index (Officer and Williamson 2012).

Table B.2: Distribution of census wealth by officeholding status.

Variable	Status	N	Min.	Mean	Max.	S.d.
<i>1820 Census</i>						
Slave wealth	Non-officeholder	28,580	0	1,166	116,040	2,908
	Officeholder	3,434	0	1,714	97,230	3,946
$p < 0.001$	all	32,014	0	1,224	116,040	3,041
<i>1850 Census</i>						
Real estate wealth	Non-officeholder	47,249	1	2,178	310,000	5,379
	Officeholder	5,745	8	2,579	103,200	5,628
$p < 0.001$	all	52,994	1	2,221	310,000	5,408

Notes: slave wealth adjusted to 1850\$ values (S. H. Williamson 2018). p -value is obtained from a Mann-Whitney-Wilcoxon test under the null hypothesis that officeholder and non-officeholder distributions are equal.

Table B.3: Record classification ensemble.

Algorithm	Parameters	MSE	Weight
Super Learner	default	0.02	-
Generalized boosted regression	default	0.02	0.05
GLM with elasticnet regularization	$\alpha = 0$	0.02	0
GLM with elasticnet regularization	$\alpha = 0.25$	0.02	0
GLM with elasticnet regularization	$\alpha = 0.5$	0.02	0
GLM with elasticnet regularization	$\alpha = 0.75$	0.02	0
GLM with elasticnet regularization	$\alpha = 1$	0.02	0.52
Neural network	default	0.13	0
Random forests	default	0.02	0.32
Random forests	# variables sampled = 1	0.04	0.09
Random forests	# variables sampled = 5	0.02	0
Random forests	# variables sampled = 10	0.03	0

Notes: cross-validated risk and weights used for each algorithm in Super Learner prediction ensemble for record classification model. MSE is the ten-fold cross-validated mean squared error for each algorithm. $Weight$ is the coefficient for the Super Learner, which is estimated using non-negative least squares based on the Lawson-Hanson algorithm. α is the elasticnet mixing parameter, where $\alpha = 0$ is the ridge penalty and $\alpha = 1$ is the lasso penalty. # variables sampled is the number of predictors sampled for splitting at each node.

Table B.4: Robustness: ITT treatment effects on officeholding.

Sample \ Specification	Officeholder + covariates	Officeholder (match prob.)	Officeholder (match prob.) + covariates
1805 winners & losers	0.003 [-0.01, 0.01]	0.003 [-0.008, 0.01]	0.003 [-0.008, 0.01]
1805 winners	-0.01 [-0.08, 0.05]	-0.01 [-0.07, 0.04]	-0.01 [-0.07, 0.04]
1807 winners	-0.006 [-0.02, 0.01]	-0.003 [-0.02, 0.01]	-0.005 [-0.02, 0.01]

Notes: *Officeholder (match prob.)* is the officeholder match probability. See notes to Table B.5.

Table B.5: Robustness: ITT treatment effects on candidacy.

Sample \ Specification	Candidate + covariates	Candidate (match prob.)	Candidate (match prob.) + covariates
1805 winners & losers	-0.001 [-0.007, 0.003]	-0.0004 [-0.001, 0.0005]	-0.0003 [-0.001, 0.0005]
1805 winners	0.02 [0.0005, 0.05]	0.003 [-0.001, 0.007]	0.003 [-0.001, 0.007]
1807 winners	0.002 [-0.003, 0.007]	-0.002 [-0.004, -0.0005]	0.0003 [-0.0005, 0.001]

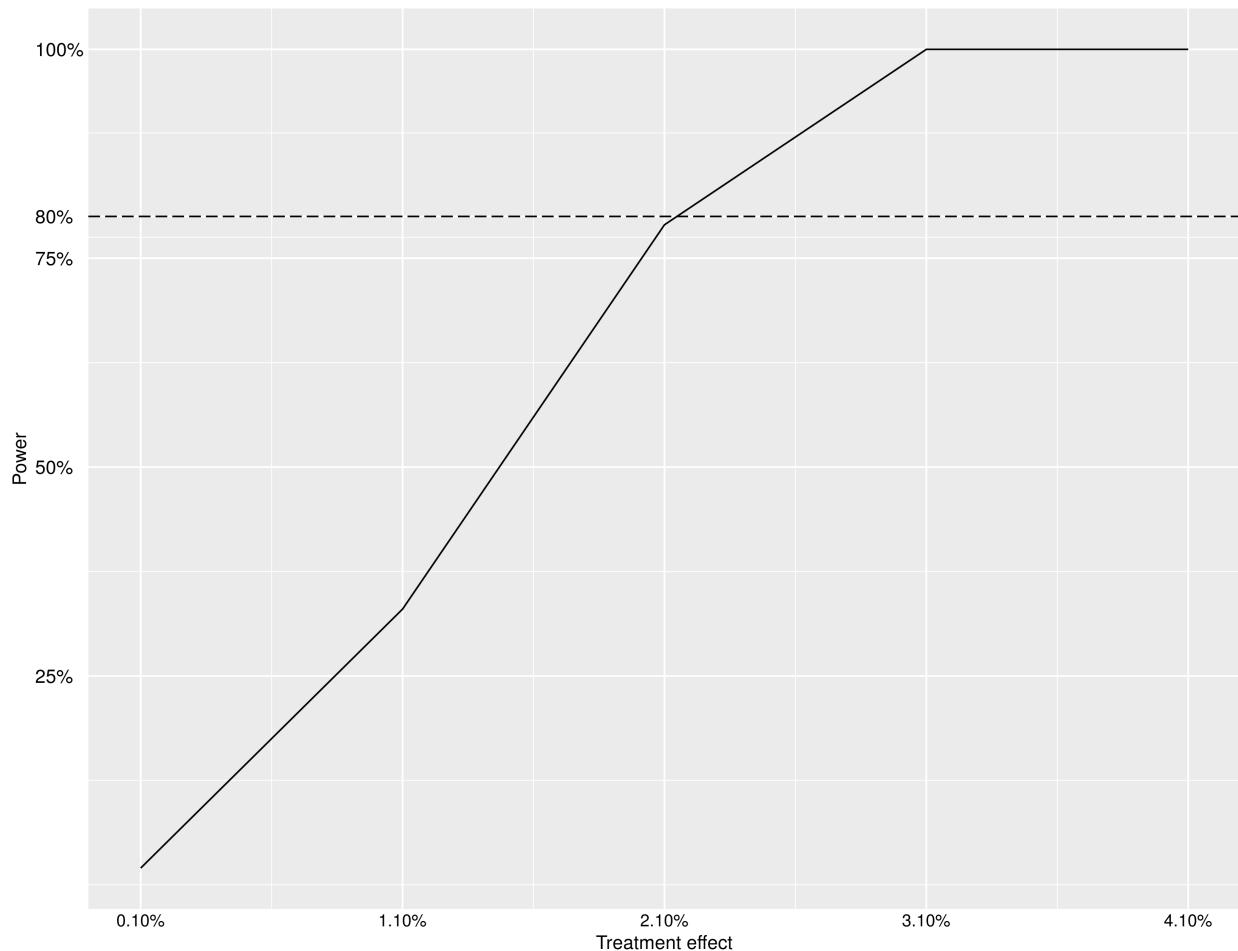
Notes: *Candidate (match prob.)* is the candidate match probability. Covariates included are those that yield $p < 0.10$ in Fig. 3.1.

Table B.6: Robustness: ITT treatment effects on slave wealth (1820\$).

Sample \ Specification	Slave wealth + covariates	Weighted slave wealth	Weighted slave wealth + covariates
1805 winners & losers	160.72 [-46.91, 368.37]	126.8 [-41.08, 294.69]	129.94 [-37.83, 297.71]
1805 winners	-194.12 [-1,294.85, 906.61]	-160.26 [-1047.19, 726.66]	-185.2 [-1075.85, 705.44]
1807 winners	-7.98 [-73.55, 57.57]	-27.2 [-80.69, 26.28]	-4.24 [-57.06, 48.58]

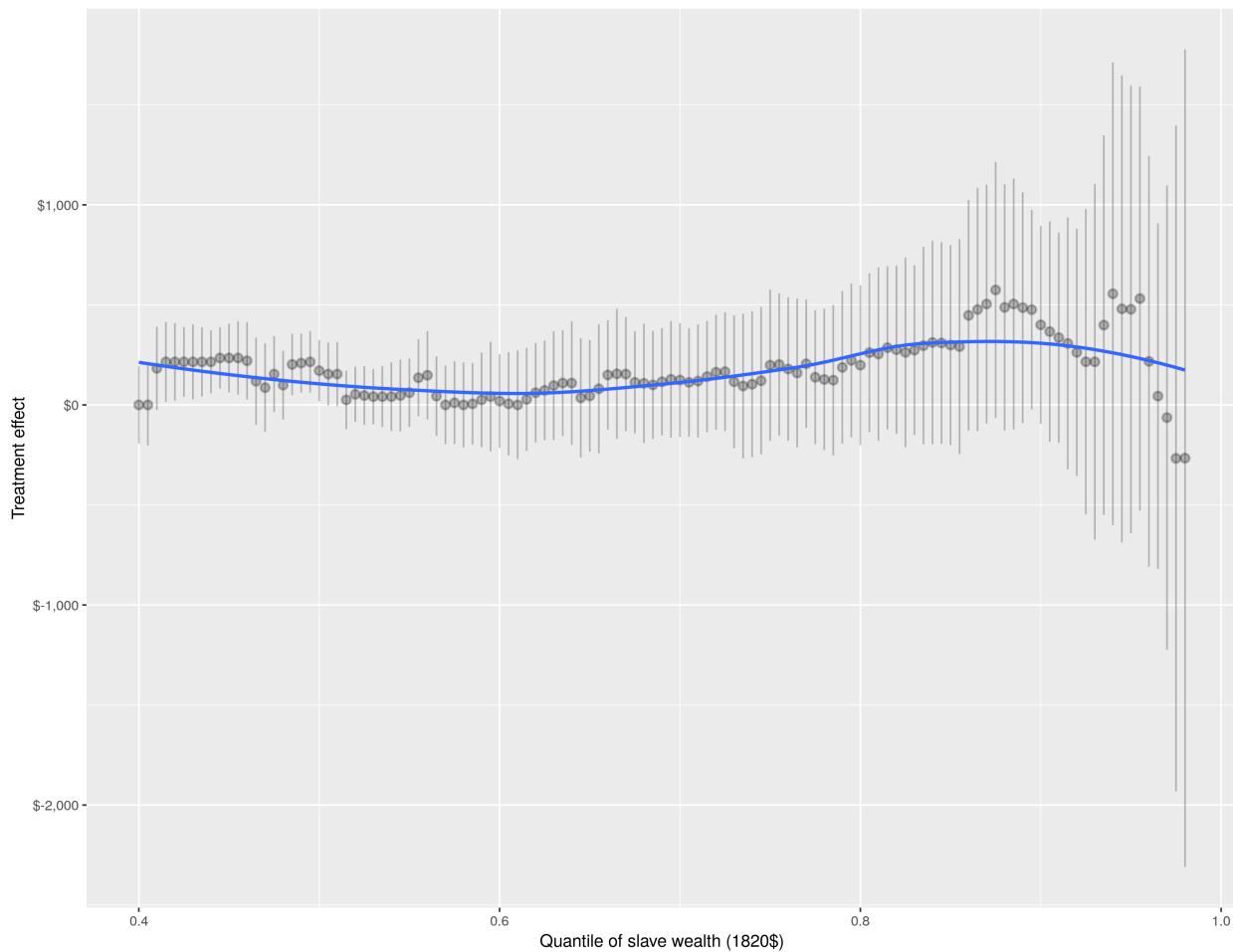
Notes: *Slave wealth (weighted)* is the same measure weighted by the census match probability. See notes to Table B.5.

Figure B.1: Power analysis by simulation for binary response variable.



Notes: $N = 21,732$ and $\mathcal{I} = 100$ iterations. The horizontal line indicates the 80% power that is normally required to justify a study.

Figure B.2: Quantile regression treatment effect estimates on slave wealth for 1805 winners & losers.



Notes: estimates from a quantile regression of the treatment effect on imputed slave wealth for participants linked to the 1820 Census ($N = 5,252$). The points are quantile-specific estimates of the treatment effect and the error bars represent 95% confidence intervals constructed from bootstrapped standard errors. Quantiles above 0.98 are omitted for display purposes. The line is a LOESS-smoothed estimate of the treatment effect.

Appendix C

Power Analysis by Simulation for Chapter 3

The purpose of a power analysis by simulation is to estimate $P(\text{Reject } H_0 | H_0 \text{ is false})$ at a fixed significance level ($\alpha = 0.05$) and sample size ($N = 21,732$) for different treatment effects $\Delta_{1,\dots,j}$. In this case N is the size of the observed sample of participants, excluding women and orphans. The simulation proceeds as follows:

1. Take a random sample of size N without replacement from the observed distribution of treatment assignments, weighted by the observed propensity score, to create a vector of simulated treatment assignments.
2. Simulate response values with Δ_j as the difference-in-means between the simulated treated and control units. Generate random values from the binomial distribution with the probability of success on each trial equal to the mean of the response in the observed sample.
3. Run linear model on the simulated data and extract the p value.

Repeat the simulation \mathcal{I} times and calculate power of the test by dividing the count of the number of p values that are less than α over \mathcal{I} . Normally, 80% power is required to justify a study. Fig. B.1 provides the results of power analysis simulations for the officeholding response.

Appendix D

Supplementary Tables and Figures for Chapter 4

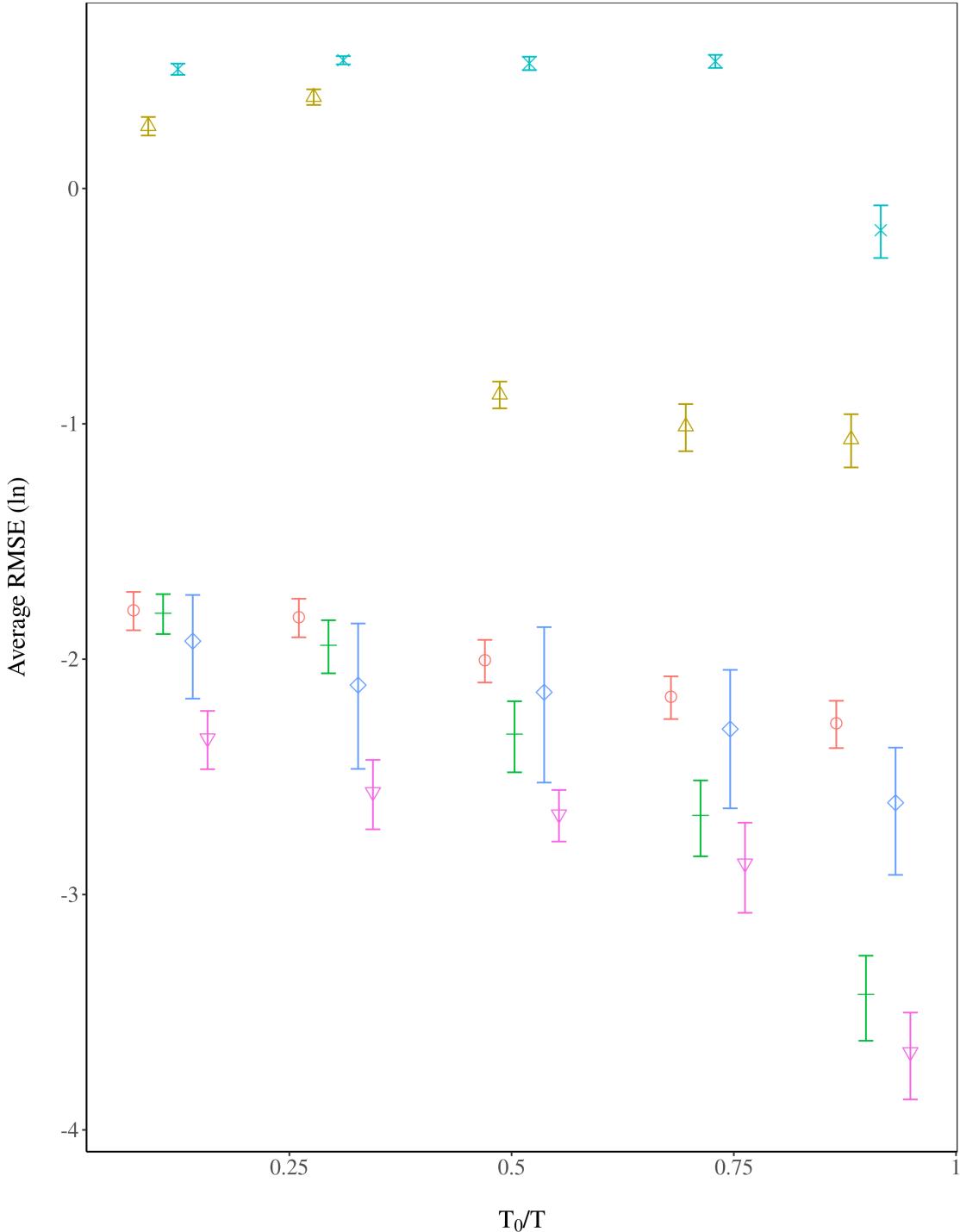


Figure D.1: Placebo tests on Basque Country terrorism data: \ominus , DID; \triangleleft , ED; $+$, MC-NNM; \times , RVAE; \diamond , SCM; ∇ , VT-EN.

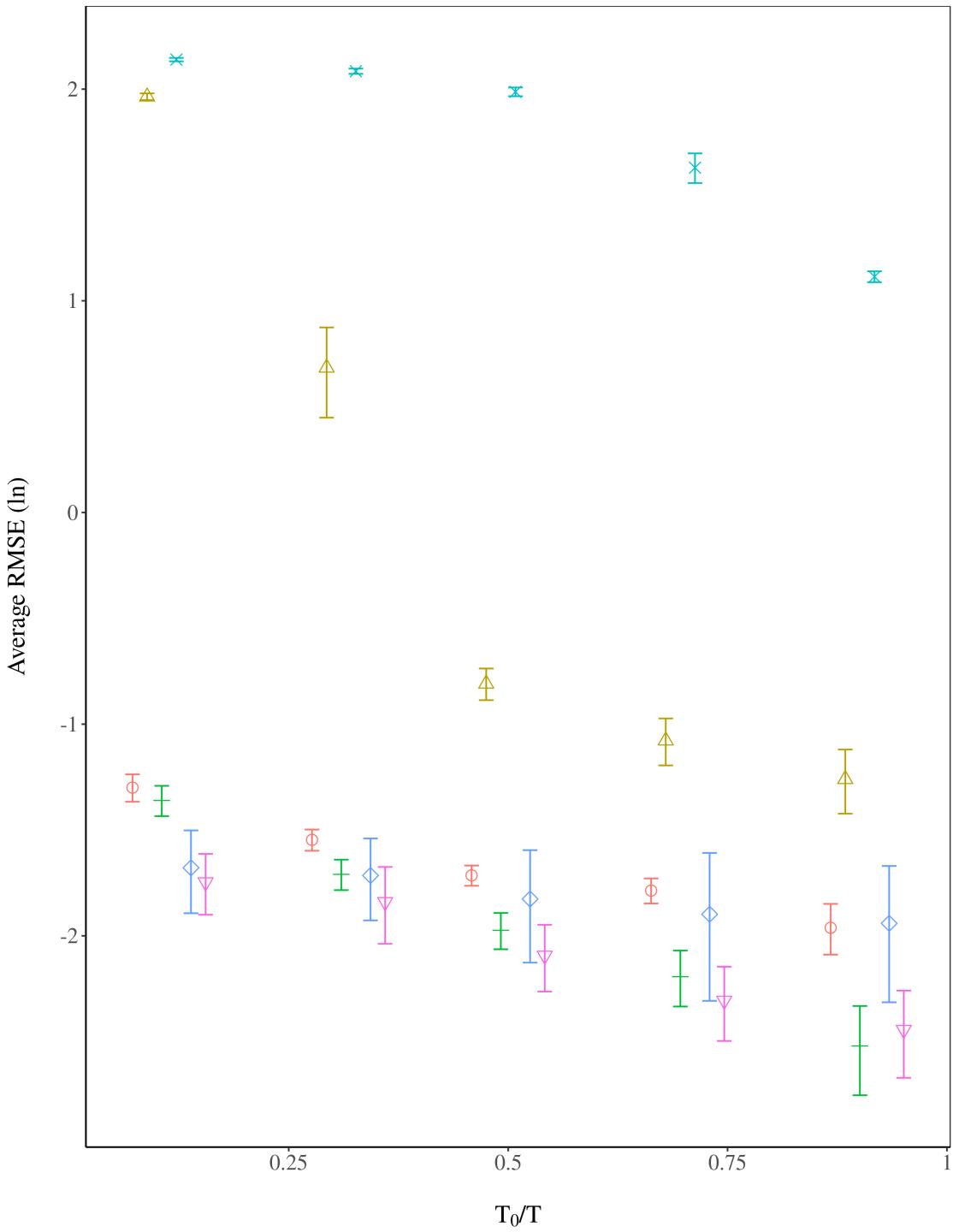


Figure D.2: Placebo tests on West German reunification data: \ominus , DID; \triangle , ED; $+$, MC-NNM; \times , RVAE; \diamond , SCM; ∇ , VT-EN.

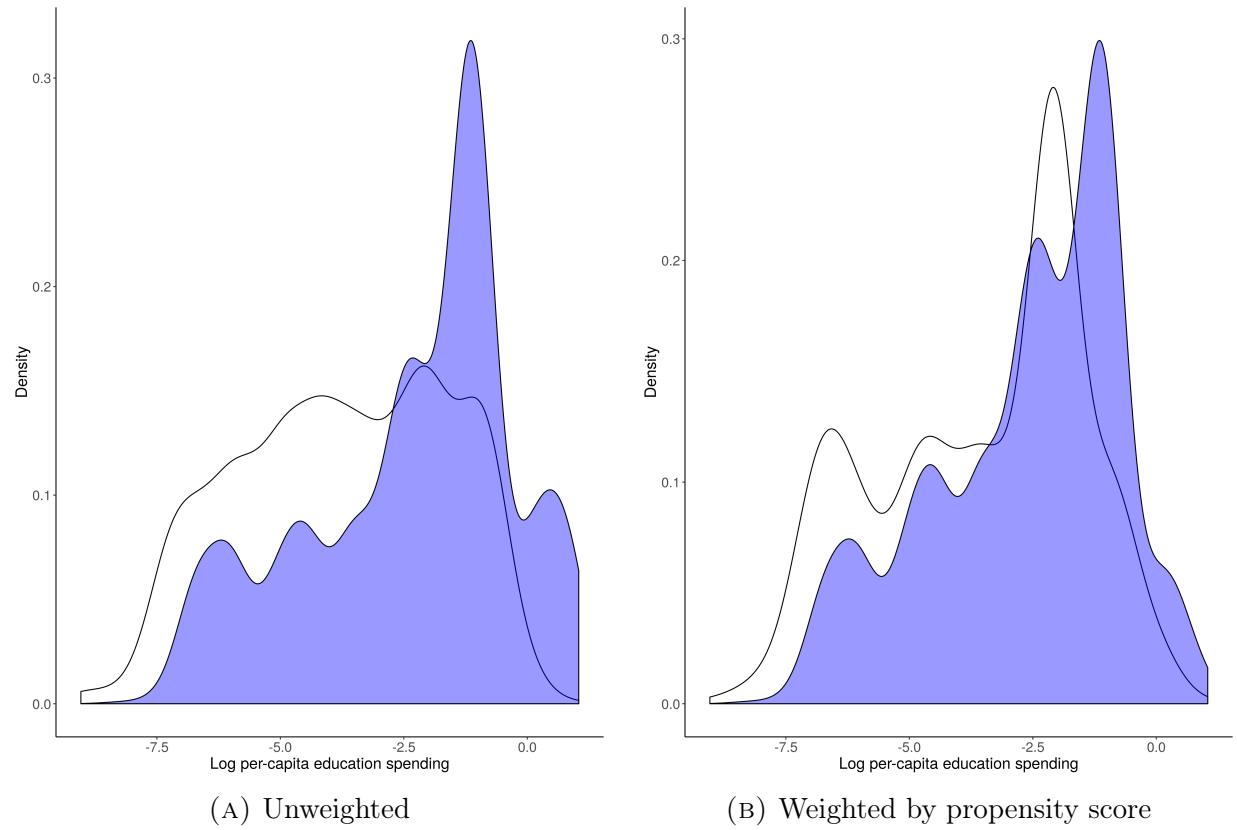


Figure D.3: Pre-period densities of log per-capita state government education spending by treatment status: \square , Control; \blacksquare , Treated

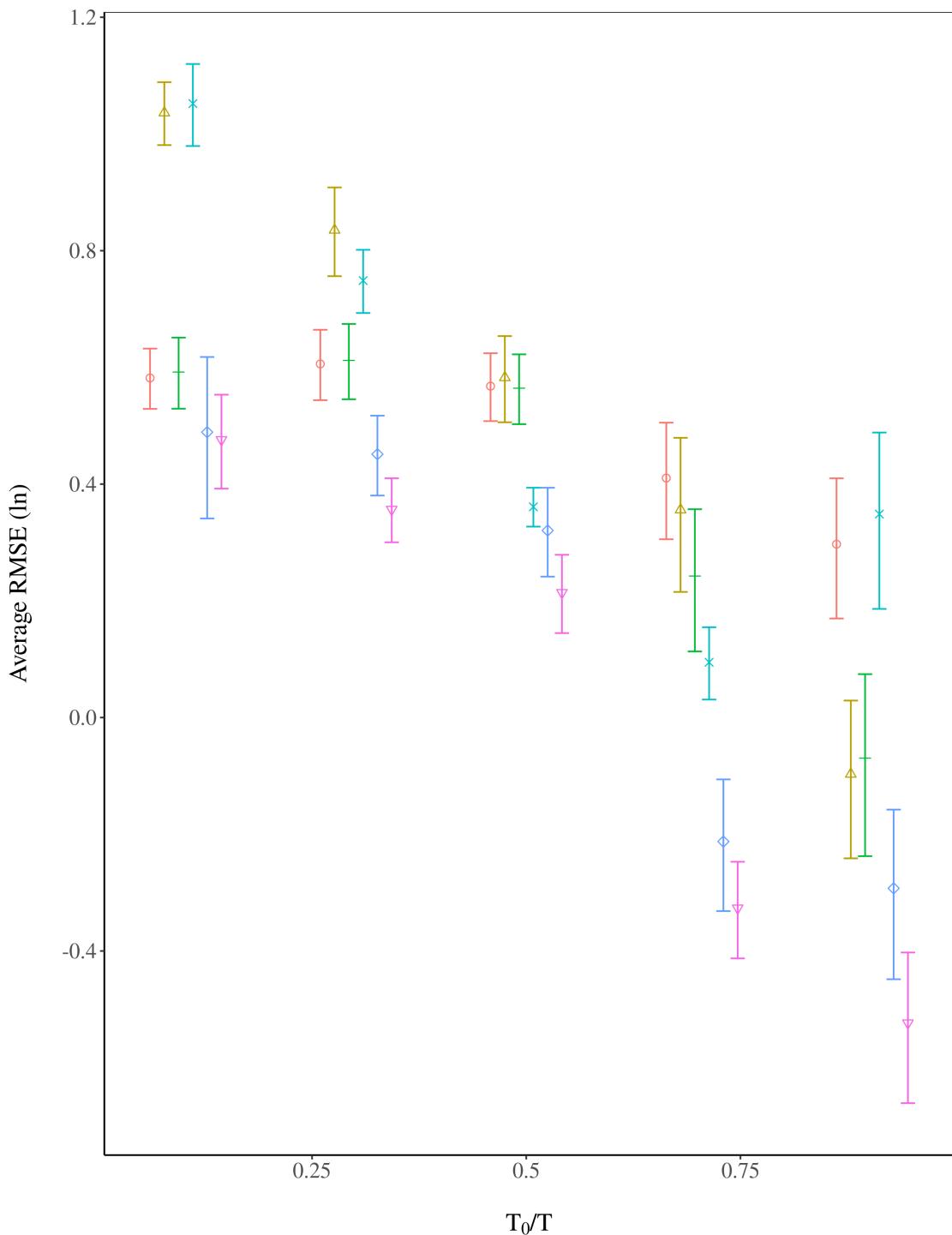


Figure D.4: Placebo tests on education spending data: \ominus , DID; \triangleleft , ED; $+$, MC-NNM; \ast , RVAE; \diamond , SCM; ∇ , VT-EN.

Appendix E

RNNs Implementation Details for Chapter 4

The networks are implemented with the `Keras` neural network library (Chollet 2015) in Python on top of a TensorFlow backend. When implementing encoder-decoder networks, the encoder takes the form of a two-layer Long Short-Term Memory (LSTM) network (Schmidhuber and Hochreiter 1997), each with 128 hidden units, and the decoder is a single-layer Gated Recurrent Unit (GRU) (Chung et al. 2014) also with 128 hidden units. Each recurrent layer uses a linear activation function (f_1) with weights initialized using Xavier initialization (Glorot and Bengio 2010). The loss function internally computes the predicted outputs as a linear function (f_2) of the log probabilities.

RNN weights are learned with mini-batch gradient descent on the WMSE using Adam stochastic optimization with the learning rate set to $5 \cdot 10^{-4}$ (Kingma and Ba 2014). As a regularization strategy, I apply dropout to the inputs and L2 regularization losses to the network weights. The networks are trained for 1,000 epochs, which takes 10 minutes to run on a laptop CPU. The model is validated on the last 20% of the training set input-out pairs.

The RVAE is implemented similarly, but with the following differences: the encoder takes the form of a single-layer LSTM with 32 hidden units and the decoder is a two-layer LSTM with the number of hidden units equal to 32 and the number of predictors, respectively. The latent space \mathbf{z} is implemented as a densely-connected layer with a dimension of 200 units and $f_3(\cdot)$ takes the form of a log-normal distribution. The RVAE is trained with stochastic gradient descent for 5,000 epochs, which takes seven minutes to run on the same CPU.

Appendix F

Hypothesis Testing for Chapter 4

Abadie, Diamond, and Hainmueller (2010) propose a randomization inference approach for calculating the exact distribution of placebo effects under the sharp null hypothesis of no effect. Cavallo et al. (2013) extends the placebo-based testing approach to the case of multiple (placebo) treated units by constructing a distribution of *average* placebo effects under the null hypothesis. Firpo and Possebom (2018) derive the conditions under which the randomization inference approach is valid from a finite sample perspective and Hahn and Shi (2017) analyze the approach from a repeated sampling perspective.

Randomization p -values are obtained following these steps:

1. Estimate the observed test static $\hat{\Phi}$ from (4.3). Averaging over the time dimension results in a T_* -length array of observed average treatment effects.
2. Calculate every possible average placebo treated effect μ by randomly sampling without replacement which $J - 1$ control units are assumed to be treated. There are $Q = \sum_{g=1}^{J-1} \binom{J}{g}$ possible average placebo effects. Since calculating Q can be computationally burdensome for relatively high values of J , I artificially set $Q = 10,000$ in cases when $J > 16$. The result is a matrix of dimension $Q \times T_*$.
3. Sum over the time dimension the number of μ that are greater than or equal to $\hat{\Phi}$.

Each element of the vector obtained from Step 3 is divided by Q to estimate a T_* -length vector of exact two-sided p values, \hat{p} .

Randomization confidence intervals

Under the assumption that treatment has a constant additive effect Δ , I construct an interval estimate for Δ by inverting the randomization test. Let δ_Δ be the test statistic calculated by subtracting all possible μ by Δ . I derive a two-sided randomization confidence interval by collecting all values of δ_Δ that yield \hat{p} values greater than or equal to significance level

$\alpha = 0.05$. I find the endpoints of the confidence interval by randomly sampling 500 values of Δ .