

Building State Capacity through Public Land Disposal: An Application of Matrix Completion for Counterfactual Prediction

Jason Poulos

University of California, Berkeley

Abstract

How would the frontier have evolved in the absence of mid-nineteenth century homestead acts? I propose using recurrent neural networks (RNNs) to predict the counterfactual time-series of frontier state capacity had there been no homestead acts. Time-specific estimates signify that homestead acts positively impacted state government finances 50 years following their implementation. Exploiting variation in the intensity of homestead entries aggregated from 1.46 million individual land patents, difference-in-differences estimates imply that homesteads significantly increased state government revenue and education spending over a period extending into the twentieth century.

Address for correspondence: Department of Political Science, 210 Barrows Hall #1950, Berkeley, CA 94720-1950. *Email:* poulos@berkeley.edu. The online Supplementary Material is available at <http://jvpoulos.github.io/papers/land-reform-sm.pdf>. I thank Sean Gailmard, Eric Schickler, Ross Mattheis, and Shom Mazumder for helpful comments. 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).

Political scientists are increasingly interested in patterns of state development across time and place. Several scholars (e.g., Bense, 1990; Murtazashvili, 2013; Frymer, 2014) theorize a relationship between mid-nineteenth century public land laws and the development of the national government. It is argued that laws designed to transfer public land to private individuals increased the bureaucratic capacity of the federal government to administer land and also reduced enforcement costs.

I argue that public land laws had long-lasting impacts on state capacity, or the ability of state governments to finance and implement policies (Besley and Persson, 2010). I explore the role of two major 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.¹ I provide evidence that homesteads authorized under these laws had long-run positive 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 laws 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 laws not only offered greater economic opportunities to eastern migrants, but also the scarcity of people on the western frontier meant that state and local governments competed with each other to attract migrants in order to lower labor costs for businesses and to increase land values and tax revenues (Engerman and Sokoloff, 2005). 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. As Engerman and Sokoloff

¹I use the terminology of “frontier” states interchangeably with “public land states” throughout the paper. Public land states are states created out of the public domain. In the South, these states include Alabama, Arkansas, Florida, Louisiana, and Mississippi. Western public land states include the 25 states that comprise the Midwestern, Southwestern, and Western U.S. (except Hawaii).

(2005) notes, the commitment of frontier state governments to offer greater political and economic opportunities for migrants was more prevalent in western states than in the South.

Garcia-Jimeno and Robinson (2009) 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 promoted equitable outcomes only when societies were initially democratic; however, when institutional quality is weak, the existence of frontier land can yield worse development outcomes because non-democratic political elites can consolidate frontier lands for themselves. Historical institutional quality as a function of fiscal capacity and newspaper entry was considerably lower in southern public land states compared to southern border states and northern states (Grosjean, 2014). Regional differences in institutional quality may explain the empirical findings of the present paper: estimates from a difference-in-differences (DD) model that leverages variation in both the timing and intensity of homesteads show that homesteads significantly lowered land inequality in western frontier counties, but had no significant impact on land inequality in southern counties.

The political economy literature is largely in agreement that inequality and state capacity are inversely related. 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. However, models that allow for differences in political influence across economic groups predict an inverse relationship. In Benabou's (2000) model, the pivotal voter is wealthier than the median and has the power to block redistribution as inequality increases. In Besley and Persson's (2009) framework, greater economic power of the ruling class reduces investment in state capacity. Landed elites might choose an inefficient organization of the state in order to create inefficiencies in

tax collection (Acemoglu et al., 2011) or “hollow-out” tax institutions in order to constrain the state’s ability to tax in the future (Suryanarayan, 2017). Similarly, Galor et al. (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).²

The paper makes a methodological contribution in applying an alternative to the synthetic control method (SCM) (Abadie et al., 2010) for inferring the effect of a policy intervention on observational time-series 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, the proposed method applies the matrix completion method (MCM) proposed by Athey et al. (2017) 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 SCM in terms of minimizing prediction error.

The paper proceeds as follows: in the section below, I overview the historical context of homestead laws and their relationship to state capacity and two potential causal mechanisms: land inequality and railroad development; Section 3 describes the proposed method of RNN-based counterfactual time-series prediction. Section ?? proposes a method of statistical inference and evaluates the proposed method against the SCM. In Section 4, I use the proposed method to estimate the long-run impacts of homestead acts on state capacity. Section 5 reports DD estimates of the effect of cumulative homesteads on state capacity, land inequality, and railroad access. Section 6 concludes.

²The inverse relationship between land inequality and state capacity has been empirically demonstrated in the context of taxes, revenues, and public school spending at the county-level in 1890 and 1930 ((Ramcharan, 2010); (Vollrath, 2013)).

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 (Lanza, 1999, pp. 13). Similar to the HSA, homesteaders could patent up to 160 acres after five years of inhabiting and improving the land, but unlike the HSA, could not commute homestead entries to cash entry after six months. Congress repealed the cash entry restriction in 1876, and sharply reversed policy in 1889 by ending cash entry in all public land states except for Missouri (Gates, 1940). In sum, the SHA followed the same application procedures as the HSA but differed in that it restricted cash sales of public land for the decade after its passage.

2.1 Land monopolization and inequality

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 HSA.³ 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 SHA. 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.

Homestead policy may have failed to create a more equitable land distribution in part due to the accumulation of public land by speculators and corporations through corrupt practices, such as the use of dummy entry-men, which is the practice of paying individuals to stake out a homestead in order to extract resources from the land with no intention of filing for the final patent. In the South, 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, pps. 216-218) argues that speculators benefited disproportionately from land laws because the economic balance of power tilted toward the wealthy.

2.2 What the railroad will bring us: More revenue

Frontier states prioritized spending on banking and transportation projects in order 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 Midwest (e.g., Illinois, Indiana, and Michigan) borrowed to invest

³Source: author's estimates using land patent data described in Section 4.

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 an 1836 act 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 et al., 2004).

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 et al. (2012) find school attendance rates increased in counties that gained access to the rail network between 1850 and 1860.

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 outcomes of affected (“treated”) units against the outcomes of unaffected (“control”) units. In the current application, public land states 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 the homestead acts — serve as control units.⁴ A common approach to the problem is the SCM,

⁴The control group includes states of the original 13 colonies, Maine, Tennessee, Texas, Vermont, and West Virginia.

which predicts the counterfactual outcomes for treated units by finding a convex combination of control units that match the treated units in term of lagged outcomes. The SCM predicts patterns across units that are assumed to remain constant over time.

This paper applies the method of matrix completion proposed by Athey et al. (2017) to predict counterfactual outcomes in a setting where multiple treated units are exposed to a binary intervention and the date of initial exposure to treatment may vary between treated units. Let \mathbf{Y} denote a $N \times T$ matrix of outcomes for each unit i, \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 first treatment exposure. Let \mathbf{Y}_{obs} define observed values of Y and \mathbf{Y}_{mis} define missing values. Define the $N \times T$ binary matrix \mathbf{M} , where $M_{it} = 1$ if Y_{it} is missing and $M_{it} = 0$ if M_{it} is nonmissing.⁵ This setup is motivated by the fundamental problem of causal inference (Holland, 1986) in that we cannot directly observe counterfactual outcomes and we instead wish to predict missing values in \mathbf{Y} for treated units with $M_{it} = 1$.

In an observational setting, units are part of the assignment mechanism that generates \mathbf{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 staggered adoption (Athey and Imbens, 2018), which differs from simultaneous adoption in that T_0 may vary across treated units. In either case, there are selection biases the missing values in \mathbf{Y} are missing not at random (MNAR) because the probability of missingness may depend on the unobserved data, $f(\mathbf{M}|\mathbf{Y}, \phi) = f(\mathbf{M}|\mathbf{Y}_{\text{mis}}, \phi)$ for all $\mathbf{Y}_{\text{mis}}, \phi$, where ϕ denotes unknown parameters (Little and Rubin, 2014, Chap. 1). Selection bias in the staggered adoption setting can occur, for example, if public land states are exposed to treatment (i.e., settled homesteads) earlier because they have higher quality land. Selection bias can occur in the simultaneous adoption setting due to latent characteristics of public land states that may be correlated with

⁵The process that generates \mathbf{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, Chap. 1).

their outcomes. The goal is to accurately estimate the effect of a policy intervention despite incomplete and biased data due to selection bias.

The connection between the matrix completion problem and the problem of inferring the effect of a policy intervention on observational data has previously been explored for rating prediction under selection bias (Schnabel et al., 2016) and in the context of domain adaptation and covariate shift, where training and test data are drawn from different distributions (Cortes et al., 2008; Bickel et al., 2009). In both problems, weighting approaches are used to solve for selection bias.

3.1 Matrix completion and imputation methods

Imputation and matrix completion are two widely-used methods for reconstructing missing data. Imputation methods reconstruct missing data by capturing synchronous relationships among observed values at the same time across different outcomes but ignore the temporal relationships within outcomes. Commonly used imputation methods include prediction model imputation, which replaces missing values with the predicted values from a regression of \mathbf{Y}_{mis} on \mathbf{Y}_{obs} . L1-regularized linear models (i.e., Lasso regression) (Tibshirani et al., 2012) are particularly useful for settings where \mathbf{Y} is a fat matrix with $N \ll T$ due to its variable selection property: it tends to shrink all but one of the coefficients of correlated outcomes to zero. These methods rely on modeling assumptions about the nature of the data-generating process and assume that the missing values in \mathbf{Y} are missing at random (MAR), or $f(\mathbf{M}|\mathbf{Y}, \phi) = f(\mathbf{M}|\mathbf{Y}_{\text{obs}}, \phi)$ for all $\mathbf{Y}_{\text{mis}}, \phi$. The MAR assumption mitigates the identifiability problems of MNAR because the probability of missingness depends only on the observed data (Tsiatis, 2007, Chap. 6). However, the MAR assumption is easily violated when considering the problem of counterfactual prediction due to the dependency structure in \mathbf{M} .

Matrix completion methods (e.g., Candès and Recht, 2009; Candès and Plan, 2010; Mazumder et al., 2010; Cai et al., 2010; Recht, 2011; Chatterjee et al., 2015; Hastie et al.,

2015) exploit relationships within and across outcomes but treat \mathbf{Y} as static and thus ignores the temporal component of the data. Matrix completion methods also typically assume the missing values in \mathbf{Y} are MAR and sampled uniformly at random.⁶ The method of matrix completion via nuclear norm minimization (NNM) proposed by Athey et al. (2017) allows for dependency structures in \mathbf{M} that arise from simultaneous or staggered adoption.⁷

3.2 Inference

The counterfactual time-series of public land states can simply be modeled as an autoregressive process in which $\mathbf{y}^{(t)}$ is linearly related to values at p previous time-steps and a series of imperfectly predictable shocks (pp. 246 ?):

$$(\mathbf{y}^{(t)} | \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(t-1)}, \theta) \sim \mathcal{N}(\alpha + \beta_1 \mathbf{y}^{(t-1)} + \dots + \beta_p \mathbf{y}^{(t-p)}, \sigma^2), \quad (1)$$

where θ is a set of estimable parameters consisting of constant term α , regression coefficients $\beta_1, \beta_2, \dots, \beta_p$, and error variance σ^2 . Least squares estimates of the parameters can be recovered by regressing

$$\mathbf{y}^{(t)} = \mathbf{x}^{(t)} = (\mathbf{y}^{(t-1)}, \mathbf{y}^{(t-2)}, \dots, \mathbf{y}^{(t-p)}) \quad \forall t = p, p+1, \dots, \tau-1. \quad (2)$$

The estimated parameters are then used to predict the counterfactual time-series of treated unit outcomes:

$$\hat{\mathbf{y}}^{(t)} = \mathbf{x}^{(t)} \quad \forall t = n+1, \dots, \tau, \quad (3)$$

where $n < \tau$ denotes the last time-step prior to a discrete intervention. The key assump-

⁶Incorrect assumptions about missing data can yield biased predictions in probabilistic models like regularized matrix factorization (Marlin and Zemel, 2009).

⁷Athey et al. (2017) remark that their estimator penalize the error terms for each unit-time observation equally without regard to the fact that the probability of missingness (i.e., the propensity score), $\Pr(M_{it} = 1 | \mathbf{Y}_{\text{obs}}, \phi)$, increases with t . The authors show that the estimator can be modified by using a loss function weighted by the propensity score, which is a solution akin to the importance weighting scheme proposed by Cortes et al. (2008) for the problem of covariate shift.

tion of this approach is that the relationship between controls $\mathbf{x}^{(t)}$ and the treated time-series $\mathbf{y}^{(t)}$ modeled prior to the intervention persists after the intervention. Under this assumption, the inferred causal effect of the intervention on the treated group is the difference between the observed time-series of the treated units and the counterfactual time-series that would have been observed in the absence of the intervention:

$$\hat{\phi}^{(t)} = \mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)} \quad \forall t = n + 1, \dots, \tau. \quad (4)$$

This treatment effect is calculated at every post-period time-step and is thus useful for understanding the temporal evolution of the causal effect.

When the counterfactual time-series is known, RNNs can be evaluated in terms of the MSPE between the predicted and actual post-intervention time-series among control units. Specifically, I calculate:

$$\text{MSPE} = \frac{1}{\tau - n} \sum_{n+1:\tau}^{\tau} \left(\hat{\phi}^{(t)} \right)^2, \quad (5)$$

where $\hat{\phi}^{(t)}$ is defined in Eq. 4.

Eq. 5 measures the accuracy of the estimated treatment effect but does not tell us anything about the statistical significance of the estimates. Abadie et al. (2010) propose a randomization inference approach for calculating the exact distribution of placebo effects under the null hypothesis. Following Cavallo et al. (2013), I extend this method to the case of multiple placebo treated units by constructing a distribution of *average* placebo effects according to the procedure described in SM-Section ??.

3.3 Simulations

I evaluate the proposed RNN-based approach on three datasets common to the SCM literature: Abadie and Gardeazabal’s (2003) study of the economic impact of terrorism in the Basque Country during the late 1960s ($J = 16$); Abadie et al.’s (2010) study of the effects

of a large-scale tobacco control program implemented in California in 1988 ($J = 38$); and Abadie et al.’s (2015) study of the economic impact of the 1990 German reunification on West Germany ($J = 16$). In each dataset, I remove the actual treated unit and evaluate the models on their ability to produce low error rates on control units; i.e., estimating treatment effects of zero.⁸ Table ?? reports the results of the placebo tests and shows that either encoder-decoder networks or LSTM outperform SCM on each of the three datasets in terms of having the lowest MSPE, with false positive rates comparable to SCM.

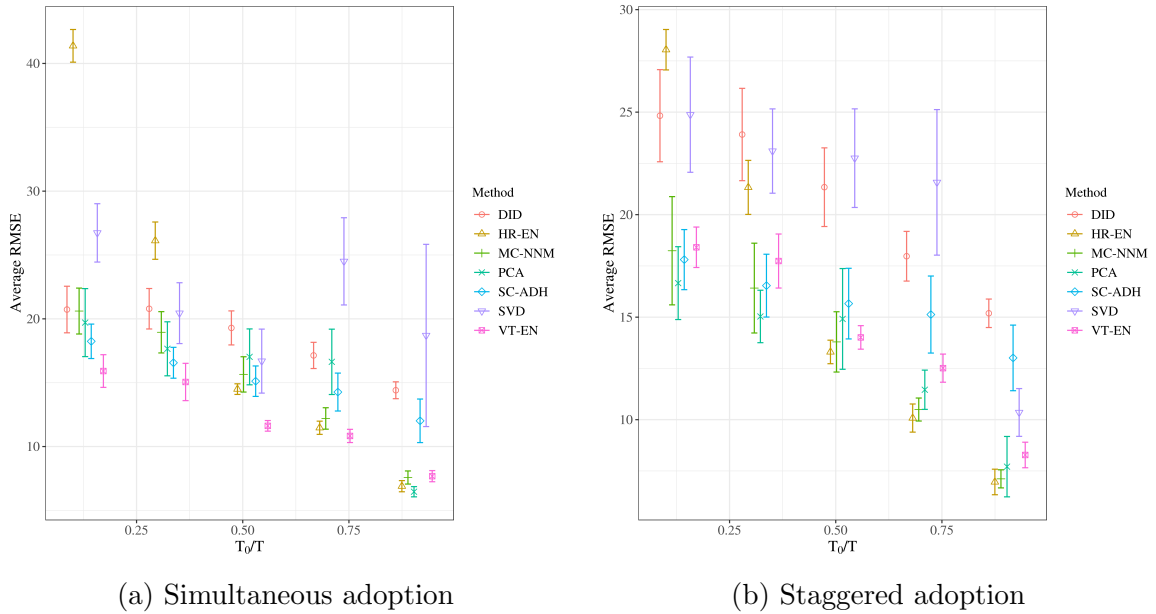


Figure 1: California Smoking Data, $N_t = 19$.

4 Impact of homestead acts on state capacity

Did homestead acts impact state capacity over the long-run? I estimate the causal impacts of homestead acts on frontier state capacity as measured by state government finances.⁹ State-level measures of total revenue and expenditure are drawn from the records of 48 state governments during the period of 1790 to 1915 and the records of 16 state governments

⁸I also calculate a single false positive rate by $FPR = \frac{FP}{(\tau-n) \times J}$, where FP is the number of false positives defined as the number of p -values less than or equal to $\alpha = 0.5$ and J is the number of placebo treated units.

⁹Variables and data sources are described in detail in Fig. SM-1.

during the period of 1933 to 1937 (Sylla et al., 1993, 1995a,b). In addition, I draw directly from U.S. census reports information on state government revenue and expenditure for the years 1902, 1913, 1932, 1942, 1962, 1972, and 1982 (Haines, 2010). I also create from the state government expenditure sub-classifications an additional state-level measure of state government education spending. I calculate per-capita measures of state government finances by taking the ratio of each measure to the total free population, which is provided in the decennial census (Haines, 2010) and take the mean of each measure across southern or western public land states to form treated unit time-series.

For each outcome, I train a separate encoder-decoder network to predict the counterfactual time-series of public land states, using only their previous history to generate predictions. Similar to the placebo tests on the SCM datasets, I evaluate the models two ways: first, I monitor the loss over 2,000 training epochs and save the model weights with the lowest error on a validation set consisting of the final 10% of the time-series.¹⁰ Second, I calculate MSPE (Eq. 5) on $J = 18$ state land states that serve as placebo treated units. Encoder-decoder networks outperform the baseline LSTM in terms of minimizing the MSPE in placebo tests (Tables SM-?? and SM-??).

Fig. ?? plots the observed and counterfactual time-series for each outcome and region. Counterfactual predictions of state government finances in the absence of homestead acts generally tracks the observed time-series until the turn of the century, at which the counterfactual flattens and diverges from the increasing observed time-series. This delay can potentially be explained by the facts that homesteaders were required to make improvements on land for five years before filing a grant and homestead entries did not substantially accumulate until after the 1889 cash-entry restriction (Figs. SM-5 and SM-7).

Taking the difference between the observed and predicted time-series (Eq. 4) yields time-specific estimates of treatment effects. Fig. ?? plots the temporal evolution of treatment effect estimates over the entire post-period and 95% randomization confidence intervals that

¹⁰The models use both dropout and L2 regularization to control for overfitting on the training set. Figs. SM-?? and SM-?? record the training history of each model.

are constructed by inverting the randomization test described in the previous section.¹¹ Fifty years after its passage, the estimated impact of the HSA on western state government education spending and revenue is 0.005 log points $[-0.16, 0.19]$, and 0.61 log points $[-0.19, 1.53]$, respectively. The confidence intervals surrounding these time-specific estimates contain zero, which implies that the estimated impacts are not significantly more extreme than the placebo treated effects estimated at the same time-step. The confidence interval on the estimated impact of the HSA on western state government expenditure in 1912, an increase of 0.17 log points $[0.004, 0.3]$, does not contain zero, which implies that the estimated impact is significantly more extreme than the placebo treated effects.

An important characteristic of the estimates plotted in Fig. ?? is the progressive widening of confidence intervals. Intuitively, counterfactual predictions become more uncertain as we move farther into the future. In the present application, confidence intervals may become implausibly wide because the post-period extends well into the twentieth century. In order to compare with the DD estimates described in the section below, I average over the entire post-period and find no evidence that the HSA impacted western state government education spending, 0.07 $[-0.32, 0.46]$, expenditure, 0.16 $[-0.21, 0.57]$, or revenue, 0.05 $[-0.33, 0.46]$. Estimates on the impact of the SHA on state capacity in the South are in the same direction and similar magnitudes.

5 DD estimates

The proposed method estimates the impact of a discrete event on a continuous outcome. However, in this application a continuous form of treatment is available: homestead entries. I derive information on homestead entries from land patent records, which provide information on the initial transfer of land titles from the federal government, including the patentee's name, issue date, number of acres, and location of the land title. The land patent

¹¹Fig. SM-?? plots the time-specific estimates of randomization p -values inferred from the exact distribution of average placebo effects under the null hypothesis.

records describe over 5 million transfers of land from the U.S. government to states, firms, and individuals. Since the records are organized by authorization, it is possible to extract information from the approximately 1.46 million records authorized under the HSA.¹²

I estimate the following DD model, where the first difference is the change in patterns of state capacity before and after implementation of the HSA or SHA, and the second difference comes from the intensity of homestead entries across frontier states:

$$y_{s,t} = \text{states}_s + \gamma d_t + X'_{s,t} \beta_t + \delta (d_t \cdot \text{homesteads}_s) + \epsilon_{s,t}. \quad (6)$$

On the left-hand-side, $y_{s,t}$ is an outcome measured in state $s = 1, \dots, N$, during period t . On the right-hand-side, states_s is a vector of states dummies to control for unobserved heterogeneity across states; d_t is a binary variable that takes the value of 1 after the enactment of the SHA in 1866 or the HSA in 1866, and 0 otherwise; and $\text{homesteads}_{s,t}$ is the log per-capita cumulative number of homestead entries up to period t . The estimated impact of the intervention is $\hat{\delta}$, which corresponds to the interaction term.¹³ $X'_{s,t}$ is the log average value of farm land, which is included to ensure that $\hat{\delta}$ is not biased by parallel trends in agricultural productivity. In theory, we would expect that homesteaders migrate to more valuable land and thus excluding average farm values from the regression would result in overestimating the actual impact of homesteads.

I apply the DD estimator on two unbalanced state-year panel datasets: the panel of southern public land states spans from 1823 to 1982 and has 38 observations prior to the enactment of the SHA. The panel of western public land states spans from 1810 to 1982 and has 52 pre-intervention years. The treatment effect estimates, summarized in Fig. 2, indicate that per-capita cumulative homesteads significantly increase per-capita revenue in

¹²Land patent records are collected by the General Land Office (GLO) and are accessible online: <https://glorerecords.blm.gov>. Land patents issued in southern public land states under the HSA authorization are presumed to be due to the SHA.

¹³I use state-stratified bootstrapped samples to construct nonparametric confidence intervals for $\hat{\delta}$. Bertrand et al. (2004) show that stratified bootstrap can be used to compute consistent standard errors when the number of groups is sufficiently large.

western states by 0.02 log points [0.0007, 0.04] and has a nonsignificant impact on revenue in southern states. Estimates on per-capita expenditure in the South and West are also nonsignificant. Per-capita cumulative homesteads significantly raise education spending in the South by 0.17 log points [0.03, 0.28], while having no impact on education spending in the West. As expected, withholding average farm values from the DD specification biases treatment effect estimates upwards (Fig. SM-2).

It should be emphasized that the DD estimator faces major disadvantages. First, the parallel trends assumption is highly restrictive and cannot be empirically verified. Second, the model assumes i.i.d. errors, which ignores the temporal aspect of the design and understates the standard errors for $\hat{\delta}$ when the regression errors are serially correlated, or $\text{Corr}(\epsilon_{s,t}, \epsilon_{s,t-1}) \neq 0$, which can arise when the time-series lengths are not sufficiently long to reliably estimate the data generating process (Bertrand et al., 2004). Lastly, the DD estimator returns only a single estimate of the treatment effect and thus does not allow researchers to chart the temporal evolution of the treatment effect.

5.1 Mechanisms: Inequality and railroads

What are the channels through which homesteads affect state capacity? Land inequality is expected to influence state capacity, although the direction of influence is theoretically ambiguous. 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).¹⁴

Table SM-2 presents DD estimates of the impact of log per-capita cumulative homesteads

¹⁴The 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. Following the procedure of Vollrath (2013), I correct for this problem by adjusting the farm Gini coefficient G by the ratio of farms to adult males, p . The adjusted coefficient is calculated as $G^A = pG + (1 - p)$.

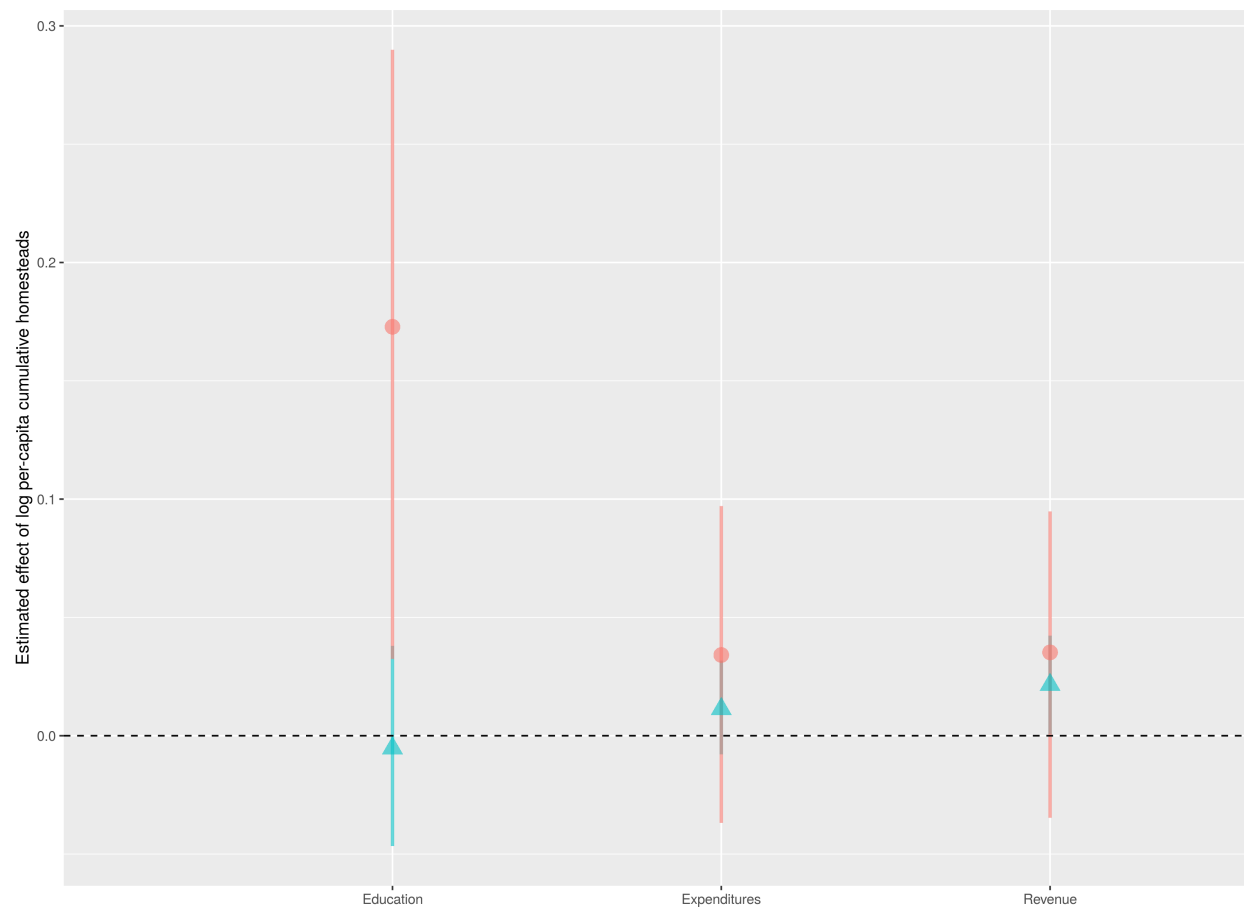


Figure 2: DD estimates of log per-capita cumulative homesteads on log per-capita state government finances. Lines represent 95% confidence intervals are constructed using 1,000 state-stratified bootstrap samples. Lines with triangles represent western public land state estimates and lines with circles represent southern public land state estimates.

on land inequality or railroad access during the period of 1860 to 1950.¹⁵ 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 western frontier state counties: a 10% increase in log per-capita cumulative homesteads is expected to lower land inequality by -0.0004 [-0.0005, -0.0002] points. The estimated impact on land inequality in the South is in the same direction, but not significant.

Railroad access is theoretically expected to increase the capacity of county and state governments by increasing the returns to farm land. I create the railroad access measure using digitized railroad maps provided by Atack (2013), which contain information on the year that each rail line was built. I overlay the railroad track map over historical county borders and calculate a binary variable indicating whether railroad access existed within a county for each year. Using these data, I estimate that 29% of counties had railroad access in 1862 and 91% had access by 1911 (Fig. SM-10).¹⁶ DD estimates of the effect of log per-capita cumulative homesteads on railroad access (value between 0 and 1) can be interpreted as follows: a 10% increase in log per-capita cumulative homesteads is expected to increase railroad access in southern counties by 0.003 [0.001, 0.005] points and by 0.009 [0.007, 0.01] points in western counties. I include farm values in the specification as a proxy for economic development, which is expected to increase state capacity and correlate with other measures of development like railroad access.

6 Conclusion

Which historical processes are responsible for present-day differences in the capacity of state governments? For example, there exists considerable variation in both the amount and

¹⁵Since railroad access is measured every year, I take the mean of railroad access to the nearest decennial year; e.g., $y_{s,1870}$ is the mean of the access measure between 1862 and 1870 in county s .

¹⁶The railroad access measure defines access with respect to county boundaries, which Atack et al. (2012) point out has limitations because a county without access might be adjacent to one with access and county boundaries frequently changed over time.

revenue sources of state and local government funding for public education: New York spent almost twice the national average per-pupil, primarily using local (54%) and state (41%) revenue sources, while Idaho spent about 60% of the national average from a combination of state (63%), local (26%) and federal (11%) sources.¹⁷

The findings of this paper signify that mid-nineteenth century homestead acts had positive impacts on frontier state government finances that can help explain contemporary differences in state capacity. RNN estimates imply that the HSA had a significant and positive impact on western state government expenditure about 50 years following its implementation. The delayed impacts can be explained by the facts that settlers were required to make improvements on land for five years before filing a grant and also homesteads did not substantially accumulate until after the 1889 cash-entry restriction. I find no evidence that the homestead acts had a significant impact on the state capacity of frontier states on average over the entire-post period that extends into the twentieth century. The inability to identify a significant average impact can be attributed to the progressive widening of confidence intervals over the post-period: the uncertainty of making counterfactual predictions based on the previous histories of (placebo) treated units increases as we move farther from the intervention year.

I also estimate a DD model that leverages variation in both the timing and the intensity of cumulative homesteads across public land states and find significant positive impacts on state capacity. I include in the DD specification average farm values in order to control for homesteaders seeking more productive lands. The DD estimates imply that per-capita cumulative homesteads significantly increase per-capita revenue in western states by 0.02 log points and raise education spending in the South by 0.17 log points. The DD estimates are of similar magnitude and direction than the RNN estimates averaged over the post-period, although the confidence intervals around the DD estimates are considerably more narrow and possibly overoptimistic due to serial correlation in the DD regression errors.

¹⁷Source: 2014 Annual Survey of School System Finances, U.S. Census Bureau. <https://www.census.gov/programs-surveys/school-finance.html>.

I explore land inequality and railroad access as possible causal mechanisms underlying the relationship between land reform and state capacity. DD estimates reveal that per-capita cumulative homesteads lowered land inequality in western counties, but did not significantly alter the distribution of land ownership in southern counties. Railroad access is theoretically expected to expand the property tax bases of state governments by increasing the returns to agricultural land. I find that cumulative homesteads significantly increased railroad access in frontier counties over a period extending into the twentieth century.

This paper makes a methodological contribution in proposing a novel alternative to SCM for estimating the effect of a policy intervention on an outcome over time in settings where appropriate control units are unavailable. In placebo tests, RNN-based models outperform SCM in terms of predictive accuracy while yielding a comparable proportion of false positives. RNNs have advantages over SCM in that they are structured for sequential data and can learn nonconvex combinations of predictors, which is useful when the data-generating process underlying the outcome depends nonlinearly on the history of predictors.

References

- Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program. *Journal of the American Statistical Association* 105(490), 493–505.
- Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative politics and the synthetic control method. *American Journal of Political Science* 59(2), 495–510.
- Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *The American Economic Review* 93(1), 113–132.
- Acemoglu, D. and J. A. Robinson (2008). Persistence of power, elites, and institutions. *American Economic Review* 98(1), 267–293.
- Acemoglu, D., D. Ticchi, and A. Vindigni (2011). Emergence and persistence of inefficient states. *Journal of the European Economic Association* 9(2), 177–208.
- Allen, D. W. (1991). Homesteading and property rights; or, “how the west was really won”. *The Journal of Law and Economics* 34(1), 1–23.
- Ansell, B. and D. J. Samuels (2015). *Inequality and Democratization: An Elite Competition Approach*. New York, NY: Cambridge University Press.
- Atack, J. (2013). On the use of geographic information systems in economic history: The american transportation revolution revisited. *The Journal of Economic History* 73(2), 313–338.
- Atack, J., R. Margo, and E. Perlman (2012). The impact of railroads on school enrollment in nineteenth century america.
- Atack, J. and R. 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(2).

- Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2017). Matrix completion methods for causal panel data models. *arXiv:1710.10251*.
- Athey, S. and G. Imbens (2018). Design-based analysis in difference-in-differences settings with staggered adoption. *arXiv preprint arXiv:1808.05293*.
- Benabou, R. (2000). Unequal societies: Income distribution and the social contract. *American Economic Review*, 96–129.
- Bensel, R. F. (1990). *Yankee Leviathan: the Origins of Central State Authority in America, 1859-1877*. Cambridge University Press.
- Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* 119(1), 249–275.
- Besley, T. and T. Persson (2009). The origins of state capacity: Property rights, taxation and politics. *American Economic Review* 99(4), 1218–1244.
- Besley, T. and T. Persson (2010). State capacity, conflict, and development. *Econometrica* 78(1), 1–34.
- Bickel, S., M. Brückner, and T. Scheffer (2009). Discriminative learning under covariate shift. *Journal of Machine Learning Research* 10(Sep), 2137–2155.
- Boix, C. (2003). *Democracy and Redistribution*. New York, NY: Cambridge University Press.
- Cai, J.-F., E. J. Candès, and Z. Shen (2010). A Singular Value Thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4), 1956–1982.
- Candès, E. J. and Y. Plan (2010). Matrix completion with noise. *Proceedings of the IEEE* 98(6), 925–936.
- Candès, E. J. and B. Recht (2009). Exact matrix completion via convex optimization. *Foundations of Computational mathematics* 9(6), 717.

- Cavallo, E., S. Galiani, I. Noy, and J. Pantano (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics* 95(5), 1549–1561.
- Chatterjee, S. et al. (2015). Matrix estimation by Universal Singular Value Thresholding. *The Annals of Statistics* 43(1), 177–214.
- Cortes, C., M. Mohri, M. Riley, and A. Rostamizadeh (2008). Sample selection bias correction theory. In *International Conference on Algorithmic Learning Theory*, pp. 38–53. Springer.
- Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A “market access” approach. *The Quarterly Journal of Economics* 131(2), 799–858.
- Doudchenko, N. and G. W. Imbens (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *arXiv:1610.07748*.
- Engerman, S. L. and K. L. Sokoloff (2005). The evolution of suffrage institutions in the new world. *The Journal of Economic History* 65(4), 891–921.
- Ferrie, J. P. (1997). Migration to the frontier in mid-nineteenth century america: A re-examination of turner’s ‘safety valve’.
- Frymer, P. (2014). ‘a rush and a push and the land is ours’: Territorial expansion, land policy, and u.s. state formation. *Perspectives on Politics* 12(1), 119.
- Galor, O., O. Moav, and D. Vollrath (2009). Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence. *The Review of Economic Studies* 76(1), 143–179.
- Garcia-Jimeno, C. and J. A. Robinson (2009). The myth of the frontier.
- Gates, P. W. (1940). Federal land policy in the south 1866-1888. *The Journal of Southern History* 6(3), 303–330.

- Gates, P. W. (1979). Federal land policies in the southern public land states. *Agricultural History* 53(1), 206–227.
- Grosjean, P. (2014). A history of violence: The culture of h=honor and homicide in the US South. *Journal of the European Economic Association* 12(5), 1285–1316.
- Haines, M. R. (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.
- Hastie, T., R. Mazumder, J. D. Lee, and R. Zadeh (2015). Matrix completion and low-rank SVD via fast alternating least squares. *The Journal of Machine Learning Research* 16(1), 3367–3402.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association* 81(396), 945–960.
- Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Lanza, M. L. (1999). *Agrarianism and Reconstruction Politics: The Southern Homestead Act*. Baton Rouge, LA: LSU Press.
- Little, R. J. and D. B. Rubin (2014). *Statistical Analysis with Missing Data*. John Wiley & Sons.
- Marlin, B. M. and R. S. Zemel (2009). Collaborative prediction and ranking with non-random missing data. In *Proceedings of the Third ACM Conference on Recommender Systems*, pp. 5–12. ACM.
- Mazumder, R., T. Hastie, and R. Tibshirani (2010). Spectral regularization algorithms for learning large incomplete matrices. *Journal of machine learning research* 11(Aug), 2287–2322.

- Meltzer, A. H. and S. F. Richard (1981). A rational theory of the size of government. *Journal of Political Economy* 89(5), 914–927.
- Murtazashvili, I. (2013). *The Political Economy of the American Frontier*. New York, NY: Cambridge University Press.
- Poulos, J. (2017, December). RNN-Based Counterfactual Time-Series Prediction. *ArXiv e-prints*.
- Ramcharan, R. (2010). Inequality and redistribution: Evidence from u.s. counties and states, 1890–1930. *The Review of Economics and Statistics* 92(4), 729–744.
- Recht, B. (2011). A simpler approach to matrix completion. *Journal of Machine Learning Research* 12(Dec), 3413–3430.
- Schnabel, T., A. Swaminathan, A. Singh, N. Chandak, and T. Joachims (2016). Recommendations as treatments: Debiasing learning and evaluation. *arXiv preprint arXiv:1602.05352*.
- Suryanarayan, P. (2017). Hollowing out the state: Franchise expansion and fiscal capacity in colonial india.
- Sylla, R. and J. J. Wallis (1998). The anatomy of sovereign debt crises: Lessons from the american state defaults of the 1840s. *Japan and the World Economy* 10(3), 267–293.
- Sylla, R. E., J. B. Legler, and J. 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.
- Sylla, R. E., J. B. Legler, and J. Wallis (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.

- Sylla, R. E., J. B. Legler, and J. Wallis (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>.
- Tibshirani, R., J. Bien, J. Friedman, T. Hastie, N. Simon, J. Taylor, and R. J. Tibshirani (2012). Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74(2), 245–266.
- Tsiatis, A. (2007). *Semiparametric Theory and Missing Data*. Springer Science & Business Media.
- Turner, F. J. (1956). *The Significance of the Frontier in American History*. Ithaca, NY: Cornell University Press.
- Vollrath, D. (2013). Inequality and school funding in the rural united states, 1890. *Explorations in Economic History* 50(2), 267–284.
- Wallis, J. J., R. E. Sylla, and A. Grinath III (2004). Sovereign debt and repudiation: The emerging-market debt crisis in the us states, 1839-1843.
- Ziblatt, D. (2008). Does landholding inequality block democratization?: A test of the “bread and democracy” thesis and the case of prussia. *World Politics* 60(4), 610–641.