Building State Capacity through Public Land Disposal: An Application of RNN-Based 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., Bensel 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.

I explore land inequality as a mechanism underlying the positive relationship between homesteads and state capacity. 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, Ticchi, and Vindigni 2011) or "hollow-out" tax institutions in order to constrain the state's ability to tax in the future (Suryanarayan 2017). In contrast, 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 direction in which homesteads are expected to affect inequality is also ambiguous. Garcia-Jimeno and Robinson (2009) argue that frontier expansion promotes equitable outcomes when societies are initially democratic; however, when institutional quality is weak, the ruling elites can

1. I 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).

exploit public land laws to concentrate their economic power. In theory, homesteading laws were designed to offer greater economic opportunities to yeoman farmers, many of whom were migrants from eastern states where market capitalism had already developed (Kulikoff 1992, Ch. 2). In practice, public land laws were exploited by land speculators, ranchers, miners, and loggers, to accumulate public land and extract natural resources during the early stages of capitalist development (Murtazashvili 2013, pp. 217). Controlling for homesteaders selecting into more productive lands, a difference-in-differences (DD) model that leverages variation in both the timing and intensity of homesteads estimates that homesteads significantly lowered land inequality in western frontier counties, but had no significant impact on land inequality in southern counties.

The paper makes a methodological contribution in proposing an alternative to the synthetic control method (SCM) (Abadie, Diamond, and Hainmueller 2010) for inferring the effect of a policy intervention on observational time-series data. Building on a new literature that uses machine learning techniques such as matrix completion (Athey et al. 2017) and Lasso (Doudchenko and Imbens 2016) for counterfactual prediction, the proposed method uses recurrent neural networks (RNNs) to predict the treated unit time-series in the absence of the intervention. I perform placebo tests on datasets introduced by the SCM literature and find that RNNs outperform SCM in terms of predictive accuracy while yielding a comparable proportion of false positives. RNNs are structured to handle sequential data, can learn nonconvex combinations of predictors, and have been shown to outperform various linear models on time-series prediction tasks (Cinar et al. 2017).

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 4 proposes a method of statistical inference and evaluates the proposed method against the SCM. In Section 5, I use the proposed method to estimate the long-run impacts of homestead acts on state capacity. Section 6 reports DD estimates of the effect of cumulative homesteads on state capacity, land inequality, and railroad access. Section 7 concludes.

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.[2] 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 in canals and railroads, while those in the South (e.g, Arkansas, Florida, and Mississippi) borrowed in order to charter state banks. Indiana,

2. Source: author's estimates using land patent data described in Section 5.

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, Sylla, and Grinath III 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, Margo, and Perlman (2012) find school attendance rates increased in counties that gained access to the rail network between 1850 and 1860.

3 Estimation

An important problem in the social sciences is estimating the effect of a policy intervention on an outcome over time. When interventions take place at an aggregate level (e.g., city or state), researchers make causal inferences by comparing the post-intervention outcomes of affected ("treated") units against the outcomes of unaffected ("control") units. Control units are typically selected on the basis of measures of similarity between the pre-intervention characteristics of treated and control units.

However, appropriate controls are not always available, either because they are fundamentally different from treated units or because the intervention indirectly affects control units. States that were not crafted from the public domain and were therefore not directly affected by the homestead acts — the states of the original 13 colonies, Maine, Tennessee, Texas, Vermont, and West Virginia — are not appropriate control units because their governments and markets had already been well-developed by the time of the passage of the homestead laws. Additionally, it is likely that state land states were indirectly affected by the out-migration of homesteaders from frontier states.

Let $\mathbf{y}^{(t)} = \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(\tau)}$ denote a multivariate time-series of observed treated unit outcomes with the temporal ordering $t = 1, \ldots, \tau$. 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 Little and Rubin 2014):

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

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

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

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

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

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

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

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

3.1 RNNs

RNNs are a class of neural networks that take advantage of the sequential nature of time-series data by sharing model parameters across multiple time-steps (El Hihi and Bengio 1995). RNNs consist of a hidden state $\mathbf{h}^{(t)}$ and an output $\mathbf{y}^{(t)}$ which operate on a sequence $\mathbf{x}^{(t)}$. At each time-step t, RNNs input $\mathbf{x}^{(t)}$ and pass it to the hidden state, which is updated with a function $g^{(t)}$ using the entire history of the sequence (pp. 337 Goodfellow, Bengio, and Courville 2016):

$$\mathbf{h}^{(t)} = g^{(t)}\left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(t)}\right) = f\left(\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}\right), \tag{5}$$

where $f(\cdot)$ is a nonlinear function that operates on all time-steps and input lengths. The updated hidden layer is used to generate a sequence of output values $\mathbf{o}^{(t)}$ in the form of log probabilities that correspond to $\mathbf{x}^{(t)}$. The loss function computes $\hat{\mathbf{y}}^{(t)} = \mathrm{linear}(\mathbf{o}^{(t)})$ and compares this value to $\mathbf{y}^{(t)}$.[3]

3.2 Encoder-decoder networks

A special variant of RNNs that are suitable for handling variable-length sequential data are encoder-decoder networks (Cho et al. 2014). Encoder-decoder networks are the standard for neural machine translation (Bahdanau, Cho, and Bengio 2014; Vinyals et al. 2014) and are widely used for predictive tasks, including speech recognition (Chorowski et al. 2015) and time-series forecasting (Zhu and Laptev 2017).

Encoder-decoder networks are trained to estimate the conditional distribution of the output sequence given the past input sequence, e.g., $p(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(t)} | \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(t)})$, where the input and

3. In the applications, loss is calculated in terms of mean squared prediction error (MSPE), $L^{(t)}_{\text{MSPE}} = \mathrm{E}\left[\left(\mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)}\right)^2\right]$.

output sequence lengths can differ. The encoder RNN reads in $\mathbf{x}^{(t)}$ sequentially and the hidden state of the network updates according to Eq. 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

$$\mathbf{h}^{(t)} = f\left(\mathbf{h}^{(t-1)}, \mathbf{y}^{(t-1)}, \mathbf{c}\right), \tag{6}$$

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

$$P(\mathbf{y}^{(t)}|\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(t-1)}, \mathbf{c}) = f\left(\mathbf{h}^{(t)}, \mathbf{y}^{(t-1)}, \mathbf{c}\right). \tag{7}$$

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

4 Model evaluation and statistical inference

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:

$$\text{MSPE} = \frac{1}{\tau - n} \sum_{n+1:\tau}^{\tau} \left(\hat{\phi}^{(t)}\right)^2, \tag{8}$$

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

Eq. 8 measures the accuracy of the estimated treatment effect but does not tell us anything about the statistical significance of the estimates. Abadie, Diamond, and Hainmueller (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 1.

I evaluate the proposed RNN-based approach on three datasets associated with the following SCM papers: Abadie and Gardeazabal's (2003) study of the economic impact of terrorism in the Basque Country during the late 1960s ($J = 16$); Abadie, Diamond, and Hainmueller's (2010) study of the effects of a large-scale tobacco control program implemented in California in 1988 ($J = 38$); and Abadie, Diamond, and Hainmueller'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.[4] Table 1 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.

The baseline LSTM performs well in all applications and outperforms encoder-decoder networks in two of the datasets. When applied to datasets with low-dimensional predictor sets, the LSTM can perform well but deeper networks such as encoder-decoder networks are susceptible to overfitting. Overfitting in this case means that the networks learn dependencies on a small subset of predictors and cannot generalize well to unseen data. Overfitting occurs when training encoder-decoder networks on the Basque Country dataset (Fig. SM-1a), which is has the lowest dimensions of the three SCM datasets. Even in this case, model check-pointing is employed so that the model with the lowest error on a validation set is used to produce counterfactual time-series. In addition, I employ dropout and L2 regularization to control for overfitting on the training set.

4. I also calculate a single false positive rate by $\text{FPR} = \frac{\text{FP}}{(\tau - n) \times J}$, 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.

Table 1: Evaluation metrics on placebo tests.

Model \ Measure	MSPE			FPR		
	Basque	California	W. Germany	Basque	California	W. Germany
Autoregressive model	0.05 ± 0.03	0.04 ± 0.06	0.001 ± 0.002	0.27	0.33	0.27
Encoder-decoder	0.01 ± 0.005	0.005 ± 0.008	0.03 ± 0.02	0.29	0.38	0.24
LSTM	0.007 ± 0.004	0.008 ± 0.009	0.02 ± 0.03	0.32	0.33	0.25
SCM	0.007 ± 0.02	0.09 ± 0.23	0.06 ± 0.2	0.31	0.33	0.25

Note: Errors represent $\pm$ one standard deviation from the MSPE.

5 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.[5] 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 during the period of 1933 to 1937 (Sylla, Legler, and Wallis 1993, 1995a, 1995b). 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.

I train a separate encoder-decoder network to predict the counterfactual time-series of each outcome, using only the previous history of the treated unit 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.[6] Second, I calculate MSPE (Eq. 8) on $J = 18$ state

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

6. The models use both dropout and L2 regularization to control for overfitting on the training set. Figs. SM-5 and SM-6 record the training history of each model.

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-2 and SM-3).

Fig. 1 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-3 and SM-4).

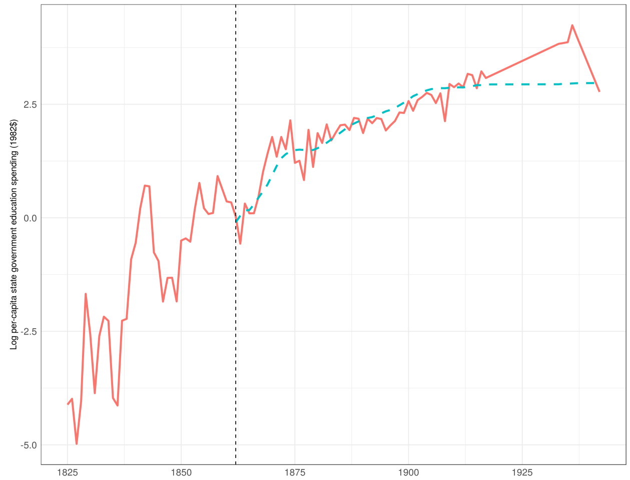
Taking the difference between the observed and predicted time-series (Eq. 4) yields time-specific estimates of treatment effects. Fig. 2 plots the temporal evolution of treatment effect estimates over the entire post-period and 95% randomization confidence intervals that are constructed by inverting the randomization test described in the previous section.[7] 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 contains 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. 2 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

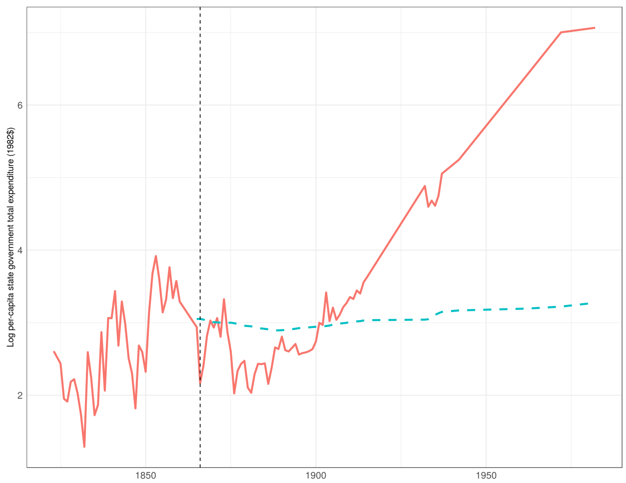
7. Fig. SM-7 plots the time-specific estimates of randomization p-values inferred from the exact distribution of average placebo effects under the null hypothesis.

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.

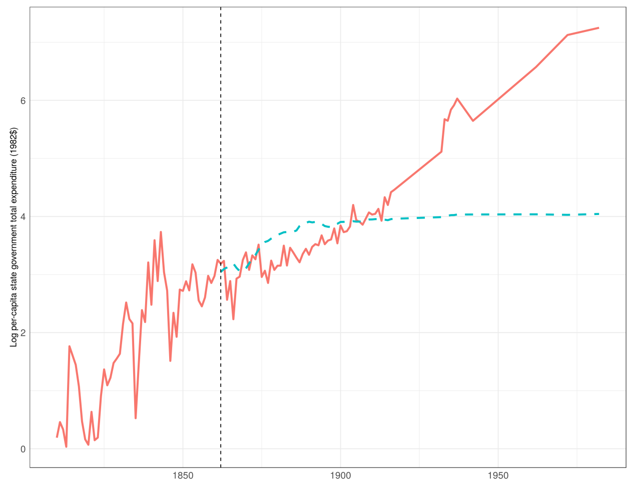
(a) Education spending in South



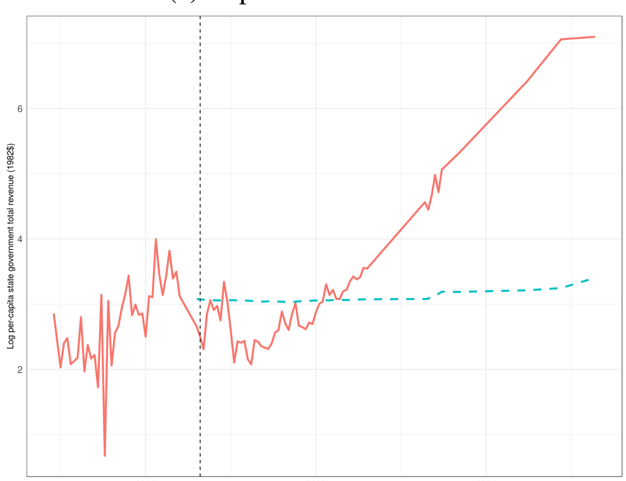
(b) Education spending in West



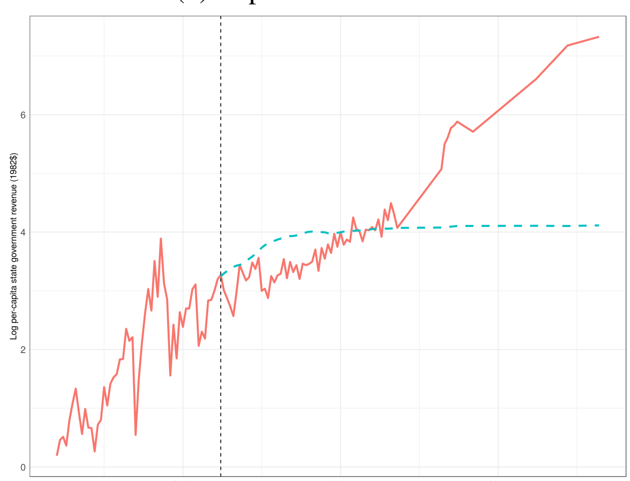
(c) Expenditure in South



(d) Expenditure in West

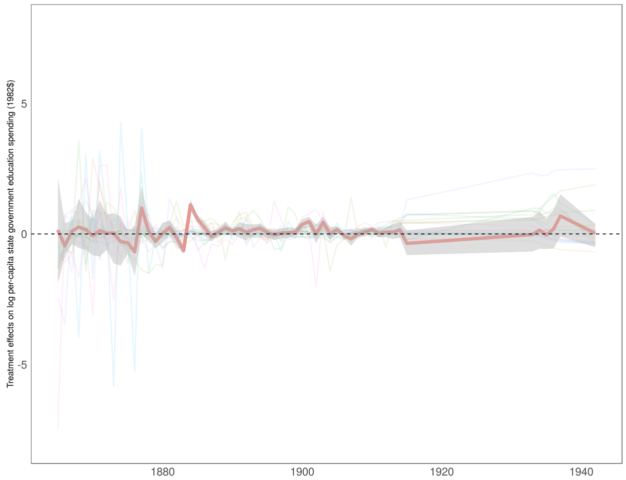


(e) Revenue in South

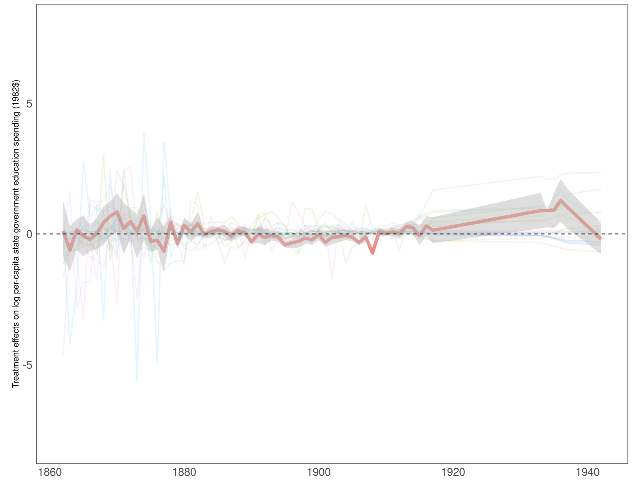


(f) Revenue in West

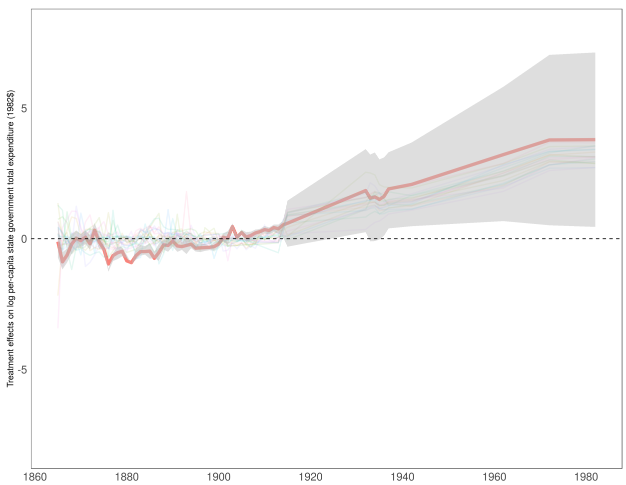
Figure 1: Observed (solid line) and counterfactual predicted (dashed line) outcomes for treated unit in state capacity datasets. Dashed vertical line represents intervention year.



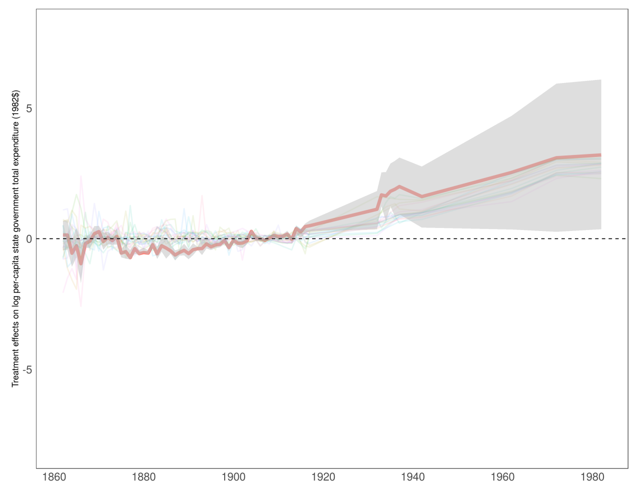
(a) Education spending in South



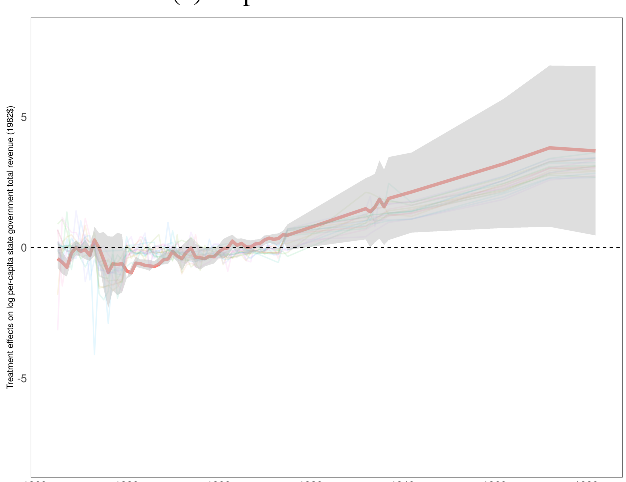
(b) Education spending in West



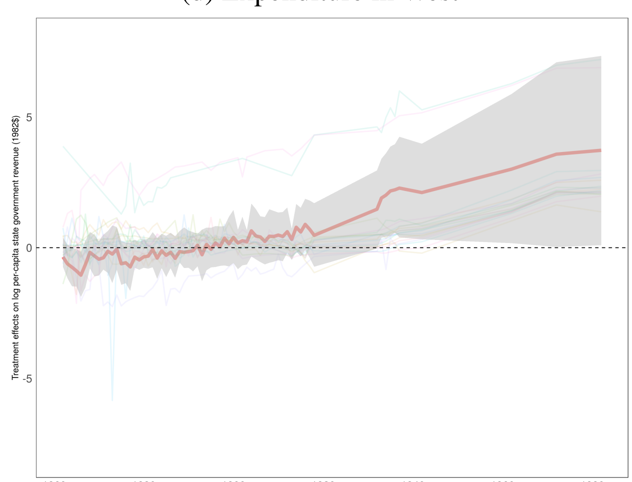
(c) Expenditure in South



(d) Expenditure in West



(e) Revenue in South



(f) Revenue in West

Figure 2: Time-series of post-period treatment effects in state capacity datasets. Darker line represents the effect on the actual treated unit and each lighter line represents the effects on placebo treated units. Shaded regions represent 95% randomization confidence intervals.

6 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 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.[8]

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:

$$y_{s,t} = \text{states}_s + \gamma\, d_t + X'_{s,t}\,\beta_t + \delta\,(d_t \cdot \text{homesteads}_s) + \epsilon_{s,t}. \tag{9}$$

On the left-hand-side, $y_{s,t}$ is an outcome measured in state $s = 1, \ldots, N$, 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.[9] $X'_{s,t}$ 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.

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

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

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. 3, indicate that per-capita cumulative homesteads significantly increase per-capita revenue in 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-3).

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 $\mathrm{Corr}(\epsilon_{s,t}, \epsilon_{s,t-1}) \neq 0$, which can arise when the time-series lengths are not sufficiently long to reliably estimate the data generating process (Bertrand, Duflo, and Mullainathan 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.

6.1 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).[10]

10. 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

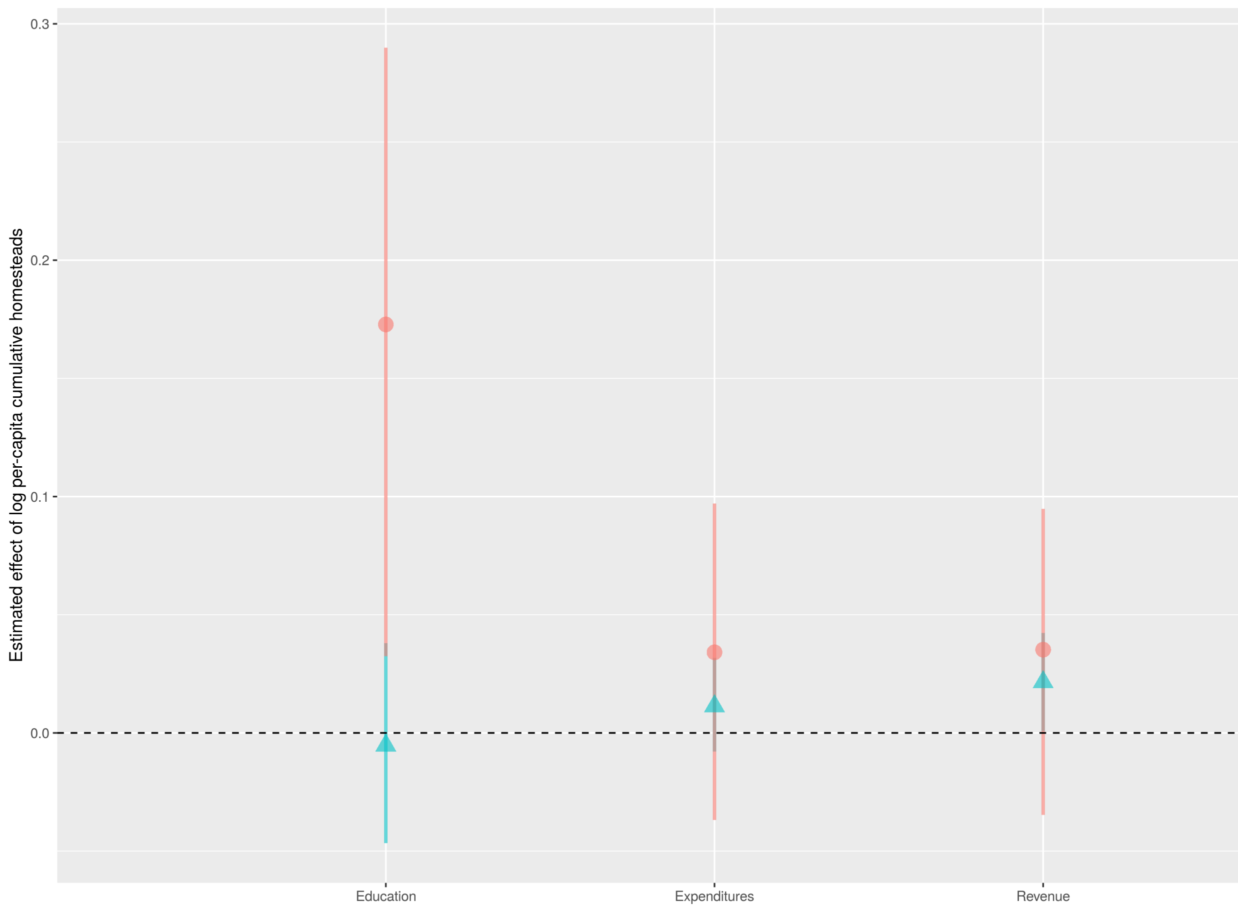


Figure 3: DD estimates of log per-capita cumulative homesteads on log per-capita state government finances. Lines represent 95% confidence intervals are constructed using 1,000 state-stratified bootstrap samples. Lines with triangles represent western public land state estimates and lines with circles represent southern public land state estimates.

Table SM-4 presents DD estimates of the impact of log per-capita cumulative homesteads on land inequality or railroad access during the period of 1860 to 1950.[11] 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-9).[12] 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.

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 $G^A = pG + (1 - p)$.

11. Since 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.

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

7 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 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.[13]

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 di-

13. Source: 2014 Annual Survey of School System Finances, U.S. Census Bureau. `https://www.census.gov/programs-surveys/school-finances.html`.

rection 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.

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 homestead entries significantly increase 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. Future work might formalize a set of assumptions necessary for the approach to be valid and provide further proof of consistency of the estimator under these assumptions.

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 (490): 493–505.

———. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59 (2): 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 (1): 113–132.

Acemoglu, Daron, Davide Ticchi, and Andrea Vindigni. 2011. "Emergence and Persistence of Inefficient States." *Journal of the European Economic Association* 9 (2): 177–208.

Allen, Douglas W. 1991. "Homesteading and Property Rights; Or, "How the West Was Really Won"." *The Journal of Law and Economics* 34 (1): 1–23.

Ansell, Ben, and David J Samuels. 2015. *Inequality and Democratization: An Elite Competition Approach.* New York, NY: Cambridge University Press.

Atack, Jeremy. 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, 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 (2).

Atack, Jeremy, Robert Margo, and Elisabeth Perlman. 2012. "The Impact of Railroads on School Enrollment in Nineteenth Century America." `http://elisabethperlman.net/papers/PerlmanMargoAtack_SchoolRR.html`.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2017. "Matrix Completion Methods for Causal Panel Data Models." *arXiv:1710.10251.*

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2014. "Neural Machine Translation by Jointly Learning to Align and Translate." *arXiv:1409.0473.*

Bensel, Richard Franklin. 1990. *Yankee Leviathan: the Origins of Central State Authority in America, 1859-1877.* Cambridge University Press.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1): 249–275.

Besley, Timothy, and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation and Politics." *American Economic Review* 99 (4): 1218–1244.

———. 2010. "State Capacity, Conflict, and Development." *Econometrica* 78 (1): 1–34.

Boix, Carles. 2003. *Democracy and Redistribution.* New York, NY: Cambridge University Press.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. "Catastrophic Natural Disasters and Economic Growth." *Review of Economics and Statistics* 95 (5): 1549–1561.

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.*

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.

Cinar, Yagmur Gizem, Hamid Mirisaee, 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.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach." *The Quarterly Journal of Economics* 131 (2): 799–858.

Doudchenko, Nikolay, and Guido W Imbens. 2016. "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis." *arXiv:1610.07748.*

El Hihi, Salah, and Yoshua Bengio. 1995. "Hierarchical Recurrent Neural Networks for Long-Term Dependencies." In *Neural Information Processing Systems,* 400:409.

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 (1): 119.

Garcia-Jimeno, Camilo, and James A Robinson. 2009. "The Myth of the Frontier." National Bureau of Economic Research.

Gates, Paul W. 1940. "Federal Land Policy in the South 1866-1888." *The Journal of Southern History* 6 (3): 303–330.

———. 1979. "Federal Land Policies in the Southern Public Land States." *Agricultural History* 53 (1): 206–227.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning.* Cambridge, MA: MIT press.

Haines, Michael 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.

Kulikoff, Allan. 1992. *The Agrarian Origins of American Capitalism.* Charlottesville, VA: University of Virginia Press.

Lanza, Michael L. 1999. *Agrarianism and Reconstruction Politics: The Southern Homestead Act.* Baton Rouge, LA: LSU Press.

Little, Roderick JA, and Donald B Rubin. 2014. *Statistical Analysis with Missing Data.* Vol. 333. John Wiley & Sons.

Mares, Isabela, and Didac Queralt. 2015. “The Non-Democratic Origins of Income Taxation.” *Comparative Political Studies* 48 (14): 1974–2009.

Murtazashvili, Ilia. 2013. *The Political Economy of the American Frontier.* New York, NY: Cambridge University Press.

Suryanarayan, Pavithra. 2017. “Hollowing Out the State: Franchise Expansion and Fiscal Capacity in Colonial India.” `https://ssrn.com/abstract=2951947`.

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`.

———. 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`.

———. 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.

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 (3): 267–293.

Vinyals, Oriol, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. 2014. “Grammar as a Foreign Language.” *arXiv:1412.7449.*

Vollrath, Dietrich. 2013. “Inequality and School Funding in the Rural United States, 1890.” *Explorations in Economic History* 50 (2): 267–284.

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.*

Zhu, Lingxue, and Nikolay Laptev. 2017. “Deep and Confident Prediction for Time Series at Uber.” *arXiv:1709.01907.*

Ziblatt, Daniel. 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.