GOVERNMENT AND ECONOMIC OUTCOMES

by

Brian Beach

B.A., The University of Washington, 2010

M.A., The University of Pittsburgh, 2011

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

THE KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Brian Beach

It was defended on

April 9th, 2015

and approved by

Werner Troesken, Professor of Economics, University of Pittsburgh

Randall Walsh, Associate Professor of Economics, University of Pittsburgh

Allison Shertzer, Assistant Professor of Economics, University of Pittsburgh

Karen Clay, Associate Professor of Economics and Public Policy, Carnegie Mellon University

Dissertation Advisor: Werner Troesken, Professor of Economics, University of Pittsburgh

GOVERNMENT AND ECONOMIC OUTCOMES

Brian Beach, PhD

University of Pittsburgh, 2015

Each chapter of this dissertation asks a specific question aimed at furthering our understanding of how a government's institutional choices and structure affect economic outcomes. I draw on modern and historical data to answer these questions as I have found that historical episodes often provide a unique opportunity to answer questions of modern relevance. This is best illustrated in the first chapter where I argue that America's 1840s state debt crisis presents a rare opportunity to study the merits of constitutional reform, a question that relates directly to ongoing debates in the Eurozone. Following the default of nine American states and territories in the early 1840s, sixteen states adopted constitutional provisions constraining their ability to borrow. These reforms helped states with tarnished reputations (i.e. defaulting states) re-establish their commitment to debt repayment. Accordingly, defaulting states were rewarded with lower borrowing costs and increased access to credit following the adoption of these reforms. In the second chapter, I study the extent to which ethnic diversity within government affects the provision of public goods. I address this question by constructing a novel dataset linking the ethnicity of California city council candidates to election outcomes and expenditure decisions. I find that higher diversity on the council leads to less spending on public goods (especially in segregated cities) and fewer votes for affected council members when they run for reelection. In the final chapter, I seek to understand how access to clean water affects human capital formation. The adoption of clean water technologies is often cited as the most important public health intervention of the twentieth century for helping eradicate typhoid fever and other waterborne diseases. I study the long-term effects of pure water by collecting annual city-level typhoid data for 75 cities, which are then merged to a unique sample linking individuals between the 1900 and 1940 censuses. Results indicate that the eradication of typhoid fever would have increased earnings in later life by one percent and increased educational attainment by one month. Put another way, the gains in earnings alone were more than sufficient to offset the cost of eradication.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 DO MARKETS REWARD CONSTITUTIONAL REFORM? LESSONS FROM AMERICA'S STATE DEBT CRISIS

America's 1840s state debt crisis presents a unique opportunity to identify whether institutional constraints lower borrowing costs. After nine states defaulted, sixteen states adopted constitutional provisions promoting credibility. Only states that defaulted during the crisis were rewarded with lower borrowing costs and increased access to credit following reform. This cannot be explained by underlying trends or differences in the content of the reforms. Non-defaulting states, which had established commitment by avoiding default, were not rewarded because reform did not convey new information. These results indicate that sovereigns with tarnished reputations can benefit from adopting constitutional constraints to convey commitment.

1.1 INTRODUCTION

The experience of Greece illustrates the importance of fiscal discipline. As a result of its recent debt crisis, GDP in Greece has fallen by 33 percent and unemployment has increased by 21 percentage-points. [1] This crisis has also made it more difficult for Greece to restructure outstanding debts and finance government activity. In April of 2010, for instance, interest rates

[1] These figures obtained by comparing statistics from the third quarter of 2008 to the first quarter of 2014. Data obtained from the National Statistical Service of Greece.

in Greece were ten percentage-points higher than in Germany.[2] But these problems are not unique to Greece. Instead, they are demonstrative of the costs sovereigns incur from entering default.[3] This paper seeks to understand how countries with tarnished reputations, like Greece, can regain favorable access to capital markets. More precisely, I ask whether financial markets reward the adoption of institutional constraints that reduce payment uncertainty.

In this paper, I use America's 1840s state debt crisis to analyze how markets respond to institutional innovations designed to promote credibility. Between 1820 and 1841, state debts increased by a factor of thirteen as states borrowed to finance canals, railroads, and banks. After experiencing a shortfall in tax revenues, states found themselves overextended, and by 1843, eight states and the territory of Florida were in default. Following these defaults, sixteen states adopted constitutional provisions that reduced payment uncertainty. To the extent that repayment could not be forced, and because most state debts were held abroad, state debts can be thought of as sovereign.[4] This presents a unique opportunity to study the response to constitutional reform using a panel of sovereigns, a setting better suited for identification.

In contrast to earlier work, which I discus below, the advantages to studying America's 1840s state debt crisis are twofold. First, because states adopted reforms at different times, and because not all states adopted reforms, it is possible to separate the response to reform from general market conditions; this also allows for the use of quasi-experimental methodology to infer causality. Second, this setting is better suited for identifying how markets respond to constitutional reform because these reforms were not threatened by wars, rebellions, or political instability. This contrasts sharply with England's reforms following the Glorious Revolution, for instance, which were threatened by the efforts to restore the Stuart monarchy (see Wells and Wills; 2000). The presence of that threat might have confounded the market response by undermining the stability of the institutional reforms.[5]

[2] Greece was effectively exiled from international capital markets in 2010. In April 2014, Greece successfully re-entered the market by issuing 3 billion euros worth of debt, but interest rates on Greek 10 year bonds remain 4.56 percentage-points higher than Germany as of June 2014. Figures obtained from Bloomberg.com.
[3] See Shambaugh, Reis, and Rey (2012) for an overview of the 2010 European sovereign debt crisis. See also Tomz and Wright (2013) and the citations within for an overview of sovereign defaults throughout history.
[4] The United States Constitution precludes suits against states to enforce payment. As a result, attempts to use the Supreme Court to compel payment have been unsuccessful. English (1996) provides a detailed discussion of the sovereignty of state debts.
[5] Consistent with this claim, Mauro, Sussman and Yafeh (2002) document sharp changes in the cost of capital resulting from wars, rebellions, and political instability.

These advantages are striking when one recognizes that previous attempts to understand the relationship between institutions and access to capital have primarily focused on a single time series.[6] North and Weingast (1989) were among the first to claim that the institutional constraints England adopted following the Glorious Revolution resulted in lower interest rates and increased access to credit. Clark (1996 and 2008), on the other hand, contends that interest rates were unaffected by these reforms while Stasavage (2002 and 2007) argues that interest rates remained high until capital owners were better represented within parliament. Wells and Wills (2000) show that financial assets responded negatively to threats to these institutions, but Sussman and Yafeh (2006) argue that those assets were weakly correlated with the cost of government debt. Sussman and Yafeh conclude that England was unable to borrow at a lower rate than other countries even though its constitution offered better protections to investors. Most recently, Cox (2012) argues that England might have been rewarded with greater access to credit even if interest rates were unaffected. Accordingly, in this paper, I analyze how constitutional reform affected both the cost of borrowing and access to credit.

Exploiting the panel feature of the state debt crisis, I analyze the market response to constitutional reform using a differences-in-differences methodology. The results indicate that defaulting states were rewarded with lower borrowing costs and increased access to credit following reform. Specifically, bonds issued by defaulting states appreciated by 13 percent following reform and outstanding debt per capita increased by $15. States that did not default on their debts, however, were not rewarded for adopting constitutional reforms. These results cannot be explained by underlying trends or by differences in the content of the new constitution. Instead, I argue that the measures states employed to avoid default demonstrated their commitment to debt repayment. Accordingly, non-defaulting states were not rewarded for adopting reforms because reform did not convey new information.[7]

These results speak directly to Title III of the European Fiscal Compact, which came into effect on January 1, 2013. The fiscal compact was designed to promote economic activity by

[6] Although this paragraph focuses on the ongoing debate regarding the Glorious Revolution, case studies from Argentina (Saiegh; 2013), Brazil (Summerhill; 2006), and Japan (Sussman and Yafeh; 2000) has also failed to bring the literature closer to a consensus.

[7] This result might explain the mixed results in the gold standard literature. Some have argued that adopting a gold standard, which serves as a commitment against inflation risk, lowers borrowing costs (Bordo and Rockoff, 1996; Obstfeldt and Taylor, 2003). More recently, Alquist and Chabot (2011) argue that the effect disappears once common risk factors are controlled for. The results in this paper indicate that whether a gold standard lowers borrowing costs might depend on underling inflation uncertainty.

fostering budgetary discipline and strengthening the coordination of economic policies in the euro area.[8] In particular, the compact requires members to adopt into domestic law balanced budget rules and procedures for reducing debt once a member's debt-to-GDP ratio exceeds 60 percent. These provisions, like those adopted in the aftermath of the 1840s state debt crisis, directly impact a sovereign's relationship with capital markets by reducing payment uncertainty. However, despite requiring these rules be adopted within the first year that the compact comes into force, several countries, including Greece, have yet to adopt reforms to comply with these requirements. The results in this paper indicate that commitment to these provisions would benefit countries whose reputation was tarnished during the recent debt crisis.

1.2 THE ORIGINS OF DEFAULT AND REFORM

1.2.a. An overview of America's 1840s debt crisis

Between 1820 and 1841, debt owed by American states increased by a factor of thirteen as states borrowed to finance canals, railroads, and banks. This era of profligate borrowing came to an end in 1841 as several states, finding themselves overextended, suspended payment on their debts. By 1843, eight states and the territory of Florida were in default. Four states eventually repaid their debts while the remaining five repudiated all or part of their debts. In response to this crisis many states adopted constitutional reforms constraining their ability to tax, borrow, and charter corporations.[9] While the decisions to default or implement reforms were not randomly assigned, it is not the case that indebtedness predicts default or that default predicts reform.

[8] The purpose of the fiscal compact is stated in Title I as: *[T]o strengthen the economic pillar of the economic and monetary union by adopting a set of rules intended to foster budgetary discipline through a fiscal compact, to strengthen the coordination of their economic policies and to improve the governance of the euro area, thereby supporting the achievement of the European Union's objectives for sustainable growth, employment, competitiveness and social cohesion.*

[9] For more information on the debt crisis see English (1996), Ratchford (1966), Wallis, Sylla, and Grinath (2004) as well as the sources within. Thomas Kettel also wrote a series of articles analyzing the debts for many states. These articles appeared in Hunt's Merchant Magazine between 1847 and 1852. For more information on constitutional reform, see Wallis (2005) and the citations within.

Table 1: Debt classifications and timeline of events

	1841 Debt per capita	Date of reform		1841 Debt per capita	Date of default	Date of resumption	Date of reform
States that did not default			**States that defaulted temporarily**				
Alabama	$26.06	-	Maryland	$32.37	1842-01	1848-01	1851-06
New York	$8.97	1846-11	Illinois	$28.42	1842-01	1846-07	1848-03
Massachusetts	$7.35	-	Pennsylvania	$19.32	1842-08	1845-02	1857-10
Ohio	$7.19	1851-06	Indiana	$18.59	1841-01	1847-07	1851-09
Wisconsin Territory	$6.45	1848-03					
South Carolina	$6.21	-	**States that partially repudiated**				
Tennessee	$4.10	-	Louisiana	$68.14	1843-02	1844	1845-05
Kentucky	$3.96	1850-06	Arkansas	$27.31	1841-07	1869-07	1846-11
Maine	$3.46	1847	Michigan	$26.47	1841-07	1846-07	1843
Virginia	$3.23	1851-10					
Missouri	$2.19	-	**States that completely repudiated**				
Georgia	$1.90	-	Florida Territory	$74.07	1841-01	-	-
Connecticut	$0.00	-	Mississippi	$18.62	1841-03	-	-
Delaware	$0.00	-					
Iowa Territory	$0.00	1846					
New Hampshire	$0.00	-					
New Jersey	$0.00	1844-09					
North Carolina	$0.00	-					
Rhode Island	$0.00	1842					
Vermont	$0.00	-					

Iowa replaced its constitution in 1846 and 1857. Louisiana replaced its constitution in 1845-05 and 1851-07. Michigan replaced its constitution in 1843 and 1850-08. Default classifications, default dates, and resumption dates obtained from Table 3 of English (1996), 1841 per capita obtained from Wallis (2005) Table 1, and type of reform obtained from Wallis (2005) Table 2. Date of reform obtained from Thorpe's Federal and State Constitutions. All information verified against Wallis's State Constitutions Database whenever possible.

The lack of correlation between indebtedness, default, and reform is illustrated in Table 1. In Table 1, states are partitioned by default status and then organized in descending order by total per capita debt in 1841. Note that Alabama did not default even though it was much more indebted than Indiana, Pennsylvania, and Mississippi. Further, Mississippi repudiated all of its debts despite being one of the least indebted defaulters – Maryland and Illinois, which were nearly twice as indebted as Mississippi, did not repudiate any of their debts. Florida, the most indebted state, defaulted but did not pursue reform, and three of the sixteen reforming states did not have any outstanding debts in 1841. So, why did states default and why did states amend or replace their constitutions? In the remainder of this section I explore the motivations behind these decisions.

Sovereign defaults typically follow declines in GDP or government revenue, and the 1840s state debt crisis is no different.[10] As land values rose during the 1830s land boom, each state saw its property tax base increase.[11] Western and southern states, in particular, anticipated a large increase in future fiscal resources because the tens of millions of acres that the federal government sold during the 1830s would finally be eligible for taxation – federal land sales were exempt from state taxation for the first five years following the sale. After borrowing against these fiscal resources, states found themselves overextended following the sharp declines in property values brought about by the Panic of 1839. As Wallis, Sylla, and Grinath (2004) argue, more established states (e.g. Massachusetts, New York, Georgia, and Alabama) were able to avoid default by quickly reinstating a property tax.[12] Pennsylvania and Maryland are an exception to this generalization. These states defaulted because they were too slow to levy adequate taxes, but both states resumed payment once taxes were in place. On the other hand, states that relied heavily on property tax revenues during the 1830s (e.g. Illinois, Indiana, Michigan, Mississippi, and Ohio) found it difficult to avoid default by raising tax rates. Consequently, many of these states were forced into default once their anticipated increase in fiscal resources failed to materialize.

Why did states reform their constitutions? As Wallis (2005) reasons, states got into trouble by using "taxless finance" to fund infrastructure investment and reform was pursued to eliminate this method of investment. Taxless finance was a unique method of funding infrastructure investment where entrepreneurs were responsible for the projects but states assumed the debt liability. Although the problems with taxless finance would become clear during the debt crisis, ex ante, taxless finance was a politically attractive way of financing infrastructure investment. The perceived benefits to taxless finance were twofold. First, because infrastructure investments were costly, taxless finance was attractive since it did not require states to raise taxes at the time of borrowing. Second, because infrastructure investments were geographically specific, taxless finance was attractive since districts that would not benefit from

[10] Tomz and Wright (2007) argue that default may be the optimal response to severe declines in exports, government revenues, or output of the tradable goods sector. Sovereigns might also find it profitable to default following a sharp increase in their cost of capital.
[11] This section is based on Wallis, Sylla, and Grinath (2004).
[12] These states largely eliminated their property tax during the 1830s, instead relying on business taxes, bank investments, and the revenues from internal improvement projects to finance expenditures. See pages 8-9 and Table 4 in Wallis et al (2004).

access to a railroad or canal would still agree to assume the project's debt liability if the expected benefit (the reduced tax burden resulting from project revenues times the probability of success) was greater than the expected cost (the increased tax burden from repaying the debt times the probability of failure). As the debt crisis unfolded, states realized that taxless finance encouraged imprudent borrowing. Accordingly, states pursued reform to prevent such a situation from occurring in the future.

The provisions states adopted are consistent with Wallis' claim that reform was pursued to prevent state and local governments from using taxless finance. For instance, section 4 of Ohio's constitution states "The credit of the State shall not, in any manner, be given or loaned to, or in aid of, any individual, association, or corporation whatever; nor shall the State ever hereafter become a joint owner, or stockholder in any company or association in this State or elsewhere, formed for any purpose whatever." This directly undermined the feasibility of taxless finance by forbidding the state from assuming the debt liability of a private entrepreneur. As another example, Illinois restricted the legislature's ability to borrow in section 37 of its constitution, which states "The State may, to meet casual deficits or failures in revenue, contract debt never to exceed in aggregate fifty thousand dollars; … and no other debt except for the purpose of repelling invasion … shall be contracted, unless the law authorizing the same shall, at a general election … receive a majority of all the votes cast." These provisions are demonstrative of the broader reforms adopted by both defaulting and non-defaulting states, reforms that constrained the legislature's ability to issue debt and charter corporations.

The types of reforms adopted were neither decided by a state's indebtedness in 1841 nor whether the state defaulted.[13] States used similar language and adopted similar provisions aimed at prohibiting the use of taxless finance. This is illustrated more precisely when one looks at the types of provisions that were adopted. Of the sixteen states that adopted reforms, thirteen adopted provisions preventing the legislature from unilaterally increasing debt, twelve imposed limits to borrowing, eight required new debt to be accompanied by a tax increase, eight prohibited the lending of credit to private individuals or corporations, ten adopted general incorporation laws, and nine prohibited the creation of corporations under special acts.

[13] In fact, the lessons of the debt crises were so salient that states joining the Union after the debt crisis (e.g. Texas, California, and Oregon) also adopted constitutions limiting the legislature's ability to borrow.

It is important to note that these provisions were innovative at the time of their adoption. To demonstrate this point I compare the text of the reformed constitutions to the previous constitutions using Wallis' state constitutions database.[14] The word "debt" appears 152 times in the reformed constitutions but only appeared eight times in the earlier constitutions. The word "corporation" appears 74 times in the new constitutions while it only appeared six times in the earlier constitutions. Lastly, the word "tax" appears 120 times in the reformed constitutions, but it only appeared 60 times in the previous constitution. These figures indicate that states adopted language constraining the state's ability to borrow, tax, and charter corporations.

Contemporary sources further validate the claim that reforms were pursued to prevent states from finding themselves in a similar situation in the future. Reflecting on the corruption and imprudent use of credit that resulted in the debt crisis, the prominent financial reporter Thomas Kettell offers his opinion on why states pursued reform. Kettell claims that "experience has brought with it the necessity of very clearly and pointedly forbidding the Legislature to exercise such powers of … grant[ing] charters, … borrowing money on their own responsibility, … [as well as] granting special privileges to corporate bodies [and] endowing them with larger credit and less liability … than is permitted to individual citizens".[15] Kettell's sentiments provide the first piece of evidence that contemporary investors viewed constitutional reforms as an effective mechanism for constraining government behavior.

1.2.b. When should we expect markets to reward constitutional reform?

The provisions states adopted dictated who had the authority to borrow, how much could be borrowed, and the purposes for which debt could be issued. This ensured that the imprudent borrowing responsible for the debt crisis was no longer possible and states would not find themselves overextended in the future. However, although these constraints serve as a credible commitment to debt repayment, whether financial markets rewarded these reforms depends on whether the reforms produced new information.

[14] This database provides the text of state constitutions and amendments through 2000 and can be accessed online: http://www.stateconstitutions.umd.edu/index.aspx. Rhode Island, Virginia, and Wisconsin were excluded from this analysis because the database is missing either the reformed constitution or the old constitution.
[15] Kettell (1851, pg. 5).

The reforms adopted by non-defaulting states should not convey new information because those states had already demonstrated commitment by avoiding default. [16] For example, New York avoided default by suspending its projects and reinstating its property tax; Alabama liquidated several branches of its state bank and reinstated its property tax; Ohio continued to finance its projects but raised property taxes dramatically – from 0.235 percent in 1837 to 0.5 percent in 1843 and 0.8 percent in 1845; and Tennessee increased its tax rate by 50 percent in order to meet its debt obligations. States incurred large costs to avoid default. But by incurring these costs, markets understood that non-defaulting states would continue to repay their future debts. As a result, states that did not default should not be rewarded for adopting reforms because the reforms did not convey new information.[17]

Whether defaulting states would continue to repay future debts was much less certain. The act of default demonstrated a willingness to impose large costs on creditors, and despite resuming payment, no state fully compensated bondholders for the losses incurred. None of these states paid interest on the missed payments, and many states adjusted the terms of repayment.[18] Further, there was not an effort to compensate the original bondholders – those that sold their claim following default. Original bondholders incurred the loss while speculators (those that purchased the bonds after default) benefited when the state resumed payment.[19] Defaulting states stood to benefit from reducing the lingering uncertainty as to whether bondholders might experience similar losses in the future. The constitutional constraints adopted not only established commitment to debt repayment but also conveyed that information to markets. Therefore, we should expect defaulting states to be rewarded for adopting constitutional reforms.

[16] Foley-Fischer and McLaughlin (2014) offer an intriguing example for how a crisis can provide the opportunity to reveal information that, in turn, reduces underlying uncertainty.

[17] This statement seemingly contradicts the findings of Dove (2012). Using a cross section of bond prices from October of 1850, 1855, and 1860, Dove documents a positive relationship between debt provisions and average bond prices, regardless of default status. The results presented in this paper suggest that the inability to separate the market response to avoiding default from the market response to adopting reform likely confounds Dove's analysis.

[18] Ratchford (1966) discusses both repudiation and debt adjustment in chapter five of his book *American state debts.*

[19] Estimates of the losses incurred by creditors when sovereigns default range from 37 to 40 percent (Benjamin and Wright, 2008; Cruces and Trebesch, 2012). Consistent with this literature, I find that bond prices fell by 38 percent following default. This figure comes from estimates of equation (1.1), presented below. The magnitude is consistent across all specifications presented in Table 3. Results available upon request.

1.3 THE MARKET RESPONSE TO REFORM

1.3.a. Data

I use data from Sylla, Wilson, and Wright's Early American Securities Database to analyze how markets responded to constitutional reform. Sylla et al gathered price quotations for publicly traded government and corporate securities between 1790 and 1860. The prices were retrieved from historical newspapers and magazines from ten cities.[20] I extract all state bond observations occurring between 1840 and 1860. This allows me to use data from each exchange to fully capture the market for state securities.[21] The frequency that observations occur varies by asset but is typically weekly, bi-monthly, or monthly. I use the average monthly price so that each asset appears at the same frequency.

These price quotations proxy for the cost of capital. Under certain arbitrage conditions, the true cost of capital for a given state equals the yield to maturity for any asset issued by that state.[22] It is not possible to analyze yield to maturity with this dataset; yield to maturity cannot be calculated without the current price, interest rate, date that coupons are issued, maturity date, and par value, and most of this information is not reported in the Sylla et al dataset. Fortunately, an asset's price and yield to maturity are inversely related. If constitutional reforms reduce payment uncertainty, then the price of assets issued by the reforming state should increase (reflecting that the asset has become less risky) and the yield to maturity will fall as a result. Therefore, price is an appropriate proxy for the cost of capital.[23] Furthermore, the magnitude of the price change is indicative of the magnitude of the change in the cost of capital.

Table 2 presents summary statistics by state. On average, the sample includes 11.5 assets for each state and each of those assets appears for an average of 43 months. Although states are not equally represented, the exhaustive list of periodicals consulted for the construction of the database suggests that these observations characterize the market for state securities between

[20] Sylla et al focused on obtaining data from financial hubs – Alexandria, VA; Baltimore, MD; Boston, MA; Charleston, SC; London, England; New Orleans, LA; New York, NY; Norfolk, VA; Philadelphia, PA; Richmond, VA.

[21] One might be concerned about the integration of capital markets during this time period. The integration of early capital markets is well documented in Neal (1990, 1992) and Sylla et al (2006). Wright (2002), in particular, shows that American markets were integrated in the antebellum period.

[22] An asset's yield to maturity is the rate of return an investor receives from holding the asset until it matures.

[23] Wells and Wills (2000) and Stasavage (2002; 2007) also use asset prices as a proxy for the cost of capital.

1840 and 1860. Arkansas, Iowa, Louisiana, and Michigan are excluded from my analysis because they appear only sporadically in the data.

Table 2: Summary statistics (1840-1860)

	Number of assets	Mean observations per asset	Median observations per asset	Total observations	Reformed constitution
States that did not default					
Alabama	5	27.2	13	136	
New York	54	19.7	14.5	1063	Y
Massachusetts	10	41.2	37	412	
Ohio	15	53.5	30	803	Y
South Carolina	3	90.3	76	271	
Tennessee	7	33.1	28	232	
Kentucky	12	43.75	29.5	525	Y
Maine	3	30	36	90	Y
Virginia	5	79.4	68	397	Y
Missouri	1	39	39	39	
Georgia	2	26.5	26.5	53	
North Carolina	1	36	36	36	
States that defaulted temporarily					
Maryland	10	69.5	37.5	695	Y
Illinois	12	31.8	18.5	381	Y
Pennsylvania	26	38.3	34	997	Y
Indiana	17	35.2	25	599	Y

Data retrieved from Early American Securities Database. See text for sample restrictions. Iowa, and Mississippi omitted because they have less than 10 total observations. Louisiana, Arkansas, and Michigan, are omitted from my analysis because they do not have enough observations following the adoption of reform.

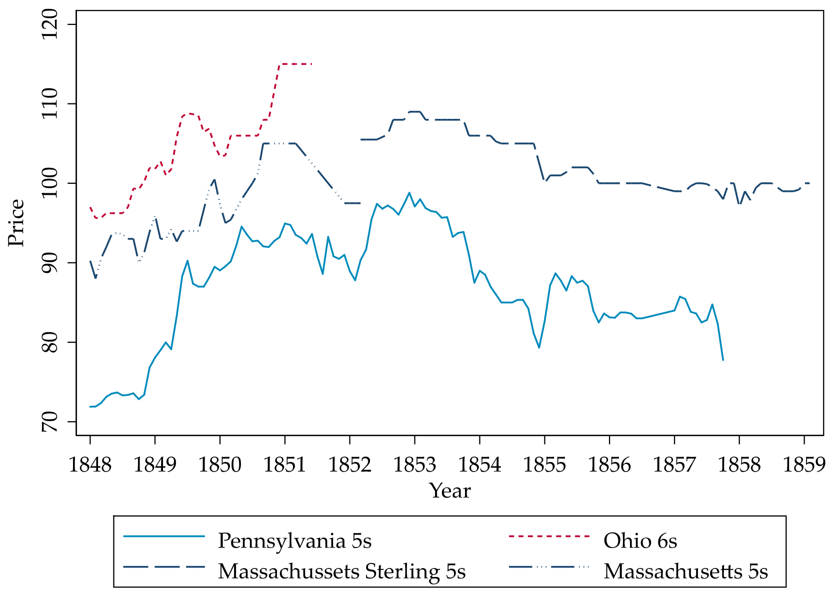
1.3.b. Methodology

I employ a difference-in-differences methodology to evaluate the market response to reform. This deviates from the cumulative abnormal return (CAR) approach that is typically used to study financial markets.[24] An abnormal return analysis involves modeling an asset's normal return, specifying an event window, and calling the sum of the residuals within that window the

[24] Campbell et al (1997) discuss the role of the abnormal market return, its history, and its applications in chapter four of their textbook *The Econometrics of Financial Markets*.

abnormal return.[25] Modeling the normal return, as is typically done in a CAR analysis, imposes a data requirement that is not easily met. In my analysis, for instance, defaulting states typically resumed payment two years before holding a constitutional convention. This leaves fewer than 24 observations to model the normal return, which is insufficient for the abnormal return methodology. Difference-in-differences, on the other hand, relaxes this data requirement.

Figure 1: Asset prices 1849-1859



Prices for Pennsylvania, a defaulting state, and Ohio, a non-defaulting state, are not plotted after they reform their constitutions. Massachusetts can be viewed as the baseline as Massachusetts neither defaulted nor reformed its constitution. Accordingly, I plot the price sequence for two bonds so that Massachusetts would appear for all years.

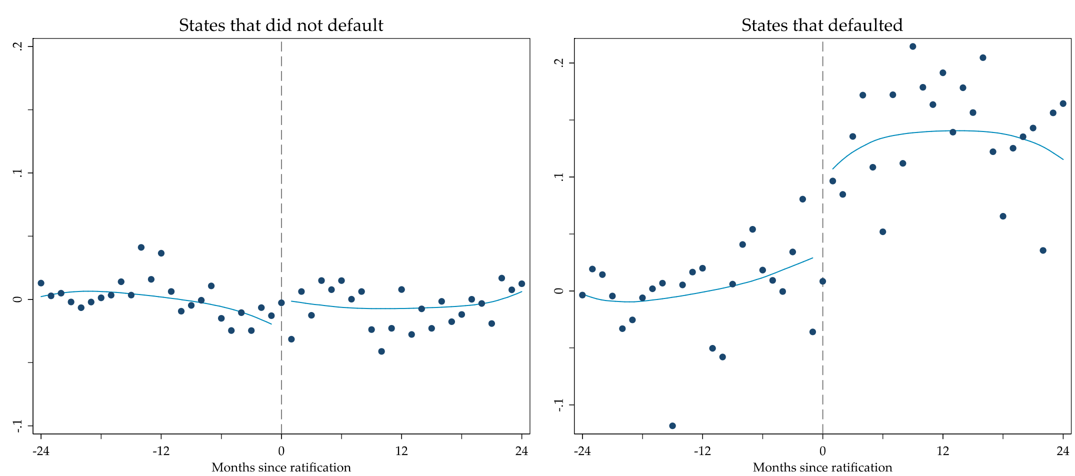
The primary advantage of difference-in-differences is that it eliminates the need to model the normal return by assuming that treated and control assets trended together prior to the treatment date. Figure 1 validates this assumption. In Figure 1, I plot the raw prices for several assets – the assets with the most observations for each group (did not default and did not reform, did not default but did reform, defaulted and reformed). I only plot one price sequence from each state because plotting raw data for several assets produces a disorderly and uninformative figure. I plot prices from 1848 until 1859. The year 1848 was chosen because it is the first year all defaulting states resumed payment and 1859 was chosen because it is one year after the last state

[25] Kothari and Warner (2006) discuss this methodology in greater detail in their handbook chapter on the econometrics of event studies.

(Pennsylvania) adopted reforms. Because the goal of Figure 1 is to illustrate that the parallel trends assumption is satisfied, I do not plot prices after reform is implemented. Figure 1 illustrates that assets trended together prior to reform.

A potential disadvantage from using a difference-in-differences approach is that, unlike a cumulative abnormal return approach, difference-in-differences imposes a functional form on the market response. I model the response as a mean shift, which assumes that markets quickly and fully capitalize the information. Figure 2 indicates that this is an appropriate way to model the market return. In this figure, I plot the average residuals for the two years before and after reform. The residuals were obtained by regressing the log of an asset's price on asset and time fixed effects.[26] Asset fixed effects normalize the price data. Time fixed effects are included because states reformed their constitutions at different times. After predicting the residuals, a non-parametric line is fitted for each regime – before reform and after reform.

Figure 2: Mean residuals at time of reform



Each point represents the mean residual across each asset and state. Residuals were obtained by regressing ln(asset prices) on time and state/asset specific fixed effects. Observations from 0 to 24 were not included in the regression so that effect of constitutional change was not captured in the asset or time fixed effects. The local polynomial smooth lines use a bandwidth of eight months.

Figure 2 is of central importance to this paper. First, Figure 2 displays my main result – that only defaulting states were rewarded for adopting reforms. This is illustrated by the 10 percent discrete jump between the two non-parametric lines in the "States that defaulted" panel

[26] When estimating the time and asset fixed effect I exclude observations occurring after constitutional reform. This ensures that the effect of constitutional reform is not captured in the fixed effects.

of Figure 2. For non-defaulting states, however, the absence of movement at the time of reform reflects the market's indifference. Second, the discrete jump for defaulting states implies that it is appropriate to model the market response as a mean shift. Although there does appear to be some movement for defaulting states in the months immediately preceding reform, this likely represents market anticipation. My preferred specification (presented in the following section) takes this into account, and as a robustness check, I implement a placebo test to illustrate that my results are not driven by underlying trends. Aside from market anticipation, the absence of a trend in the residuals prior to reform indicates that the control assets accurately model the movement of the treated assets. This provides further support for the parallel trends assumption that is necessary when using differences-in-differences.

The remainder of my analysis will study the market response to constitutional reform within a difference-in-differences framework. Specifically, I will estimate variations of the following equation:

$$\begin{aligned} P_i(t) = \alpha + \beta_1 def(t) + \beta_2 res(t) + \beta_3 con(t) * \mathbf{1}[state\ did\ not\ default] \\ + \beta_4 con(t) * \mathbf{1}[state\ defaulted] + \beta_5 post_i(t) + [asset\ FE's]_i \\ + [month\ FE's]_t + \varepsilon_i(t) \end{aligned} \tag{1.1}$$

where $P_i(t)$, denotes the log of the price of asset i during month t. The variables $def(t)$, $res(t)$, and $con(t)$ are indicator variables equal to one if asset i was issued by a state that entered default, resumed payment, or reformed its constitution by time t, respectively. I interact $con(t)$ with default status to identify whether reputation influences the market's reaction. The variable $post_i(t)$ is an asset specific indicator variable equal to one if time t occurs after the event window. This post-treatment indicator, which zeros out the effect of reform, allows me better estimate the time fixed effects by including observations outside of the event window – dropping these observations produces slightly noisier but qualitatively similar results. State fixed effects are omitted because they are captured in the assets fixed effects (assets are inherently state specific), and I include fixed effects for each month to control for general market trends. Lastly, I adjust the standard errors by clustering at the asset level.

1.3.c. Differences-in-differences results

Table 3 presents my results. The first column of Table 3 estimates a variation of equation (1.1) where the effect of reform is not allowed to vary by default status. Under this specification reform appears to elicit a positive but insignificant response from financial markets. Once the effect of reform is allowed to vary by default status, however, I find that assets issued by defaulting states appreciated by 13 percent following reform, significant at the one percent level. Assets issued by non-defaulting states, on the other hand, were unaffected by reform. These results are consistent with the hypotheses outlined in section 1.2 – constitutional reforms only convey new information for states with a tarnished reputation.[27]

Because states often held constitutional conventions to discuss the new constitution, markets might have anticipated the provisions that would be included. Although anticipation should attenuate my estimates, in column three, I capture the anticipated response by including an indicator (interacted with default status) that is equal to one if the asset was issued by a state that hosted a constitutional convention by time t. In this specification, the total market response becomes the linear combination of the convention indicator and the reform indicator. Assets issued by defaulting states increased by 18.3 percent following reform; again, non-defaulting states did not elicit a statistically significant response.

In column four, I include a time trend for defaulting states. This trend, which starts when a state resumes payment and ends with the adoption of reform, controls for the fact that markets might slowly reward defaulting states as they build a reputation for paying debts on time and in full. I find evidence that this is the case. The trend coefficient in Table 3 indicates that assets issued by defaulting states appreciated at a rate of 0.2 percent per month from the resumption of payments until the adoption of reform. The inclusion of this trend reduces the estimated effect of

[27] To deal with the possibility of serial correlation, I use the two-step procedure described in section four of Betrand, Duflo, and Mullainathan (2001). Specifically, I estimate a variation of equation (1.1) that omits both $con(t) * \mathbf{1}[state\ did\ not\ default]$ and $con(t) * \mathbf{1}[state\ defaulted]$. Then, for states that adopted reforms, I divide the residuals into state specific "pre-reform" and "post-reform" groups. Regressing those residuals on state fixed effects and the previously-omitted treatment indicators yields results similar to the results described above; assets issued by defaulting states appreciated by nearly 5% following the adoption of constitutional reforms while assets issued by non-defaulting states were unaffected by reform. This procedure, however, is not well suited for the remainder of my analysis. Because I am interested in understanding the timing of the market response, my preferred specification includes constitutional convention indicators to disentangle the anticipated response from the unanticipated response. Using the two-step procedure with these indicators yields statistically insignificant results because the treatment effect is loaded onto the convention indicator in the first stage. Accordingly, I do not use this procedure for the remainder of my analysis.

reform for defaulting states from 18 percent to 15 percent. However, if we assume that assets would have continued to increase at a rate of 0.2 percent per month throughout the event window, then the effect falls to 13.3 percent. The specification used in column four, with the assumption that assets would have continued to increase at 0.2 percent per month in the absence of reform, is my preferred specification.[28]

Table 3: Constitutional reform's effect on ln(asset prices)

	(1)	(2)	(3)	(4)
Indicator for implementing constitutional reform	0.029			
	(0.027)			
Indicator for implementing reform *(untarnished reputations)*		-0.028	-0.015	-0.014
		(0.024)	(0.016)	(0.016)
Indicator for implementing reform *(tarnished reputations)*		0.130***	0.088**	0.057
		(0.047)	(0.038)	(0.036)
Indicator for hosting a constitutional convention *(untarnished reputations)*			-0.009	0.008
			(0.030)	(0.029)
Indicator for hosting a constitutional convention *(tarnished reputations)*			0.095*	0.096*
			(0.054)	(0.051)
Monthly trend from resumption of payments until month of reform				0.002**
				(0.001)
Effect of reform for states with untarnished reputations†		-0.028	-0.025	-0.006
		(0.024)	(0.033)	(0.032)
Effect of reform for states with tarnished reputations†		0.130***	0.183***	0.133**
		(0.047)	(0.070)	(0.060)
R-squared	0.930	0.931	0.931	0.932
Observations	6729	6729	6729	6729

† The effect of reform is the linear combination of implementing reform and hosting a convention. In column (4), I subtract 12*monthly trend from the linear combination to remove the effect from the pre-existing trend.
Each regression includes an indicator for entering default, an indicator for resuming payment, and an asset specific indicator that turns on at the end of the event window (12 months after reform). Each regression also includes asset and time fixed effects. Robust standard errors, clustered at the asset-level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

[28] It is possible to incorporate a series of trends into this analysis – a trend beginning with repayment, a trend that begins at the time of reform, and a trend that begins at the end of the event window. With this approach the treatment effect becomes the differential change in trend as well as any mean shift. This produces results of similar magnitude and of the same significance. Results available upon request. Alternatively, it is possible to model the treatment effect using a series of indicators for each time period. This quasi-event-study approach produces qualitatively similar results, but depending on the size of the bins used for each indicator, the estimates can be quite noisy.

1.3.d. Are these results driven by an underlying trend?

In this section, I construct a placebo test to further illustrate that the results presented in Table 3 identify the effect of constitutional reform and not an underlying trend. I run the same specifications presented in Table 3 but instead of using the true reform and convention dates I use the date 24 months earlier. This choice ensures that the placebo test will not pick up the anticipated or unanticipated response because the event window will end before the state actually holds its constitutional convention. The results of my placebo test are presented in Table 4. The effect of constitutional reform is insignificant in each specification, which indicates that the results in Table 3 are not driven by an underlying trend.

Table 4: Placebo test - False constitutional reform's effect on ln(asset prices)

	(1)	(2)	(3)	(4)
Indicator for implementing constitutional reform	0.003 (0.033)			
Indicator for implementing reform *(states that did not default)*		-0.019 (0.031)	-0.003 (0.021)	0.004 (0.020)
Indicator for implementing reform *(states that defaulted)*		0.041 (0.063)	0.024 (0.046)	-0.001 (0.046)
Indicator for hosting a constitutional convention *(states that did not default)*			-0.020 (0.032)	-0.014 (0.032)
Indicator for hosting a constitutional convention *(states that defaulted)*			0.029 (0.038)	0.060 (0.039)
Monthly trend from resumption of payments until month of reform				0.002** (0.001)
Effect of reform for states that did not default†		-0.019 (0.031)	-0.023 (0.037)	-0.010 (0.036)
Effect of reform for states that previously defaulted†		0.041 (0.063)	0.053 (0.074)	0.035 (0.069)
R-squared	0.931	0.931	0.931	0.932
Observations	6729	6729	6729	6729

† The effect of reform is the linear combination of implementing reform and hosting a convention. In column (4), I subtract 12*monthly trend from the linear combination to remove the effect from the pre-existing trend.
Each regression includes an indicator for entering default, an indicator for resuming payment, and an asset specific indicator that turns on at the end of the event window (12 months after reform). Each regression also includes asset and time fixed effects. Robust standard errors, clustered at the asset-level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

I do not use a false date for the resumption of payments in this placebo test. This approach does not affect the estimates presented in Table 4, as the reform coefficient remains insignificant when a false resumption date is used. However, using the true resumption date provides further evidence that states were slowly rewarded for building a reputation for repaying debts in full and on time. Specifically, the estimated coefficient is statistically significant and of similar magnitude to the coefficient presented in Table 3. In both specifications, assets issued by defaulting states appreciated at a rate of 0.2 percent per month after resuming payment. This indicates that my preferred specification, where I assume that assets would have continued to increase at 0.2 percent per month in the absence of reform, is reasonable.

1.4 ROBUSTNESS CHECKS

1.4.a. Event window robustness check

In this section I ask whether the results in Table 3 are sensitive to my event window definition. Specifically, I run my preferred specification using an event window of 4, 8, 12, 16, 20, and 24 months. The results of this robustness check are presented in Table 5. The first thing to note from Table 5 is that the effect of reform for non-defaulting states is never significant. The second thing to note is that the effect for defaulting states is always positive and statistically significant. Furthermore, the effect is of similar magnitude and significance for every window. This indicates that the effect was quickly capitalized into asset values. The absence of any mean reversion indicates that the market did not overreact to reform. Moreover, the persistence of the effect suggests that the benefits from reform were not short lived.

1.4.b. Default status robustness check

The results thus far indicate that assets issued by defaulting states appreciated by 13 percent following reform. Section 1.3.d shows that the identified effect cannot be explained by an underlying trend while Section 1.4.a shows that the effect was quickly capitalized and persistent. This section analyzes the role of default status in explaining why markets only rewarded some states. To do this, I construct a placebo test where I randomly assign default status to each of the reforming states. There are 256 ways to organize the eight reforming states into two distinct groups, and for each of those combinations I run the regression specification in column three of Table 3.[29] Figure 3 plots the distribution of treatment effects for each of those regressions.

Table 5: Constitutional reform's effect on ln(asset prices) with various event windows

	4 months	8 months	12 months	16 months	20 months	24 months
Indicator for implementing reform *(states that did not default)*	-0.006	-0.009	-0.014	-0.020	-0.023	-0.016
	(0.012)	(0.014)	(0.016)	(0.017)	(0.019)	(0.016)
Indicator for implementing reform *(states that defaulted)*	0.035	0.039	0.057	0.062*	0.059*	0.059*
	(0.037)	(0.035)	(0.036)	(0.035)	(0.033)	(0.032)
Indicator for hosting a constitutional convention *(states that did not default)*	-0.000	0.008	0.008	0.008	0.006	0.006
	(0.031)	(0.030)	(0.029)	(0.029)	(0.029)	(0.028)
Indicator for hosting a constitutional convention *(states that defaulted)*	0.087*	0.098**	0.096*	0.095*	0.093*	0.101*
	(0.049)	(0.049)	(0.051)	(0.053)	(0.054)	(0.054)
Monthly trend from resumption of payments until month of reform	0.002**	0.002**	0.002**	0.002**	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Effect of reform for states that did not default†	-0.006	-0.001	-0.006	-0.012	-0.017	-0.010
	(0.034)	(0.032)	(0.032)	(0.032)	(0.032)	(0.031)
Effect of reform for states that previously defaulted†	0.114**	0.123**	0.133**	0.131**	0.118*	0.118*
	(0.055)	(0.054)	(0.060)	(0.063)	(0.064)	(0.064)
R-squared	0.932	0.933	0.932	0.932	0.931	0.932
Observations	6729	6729	6729	6729	6729	6729

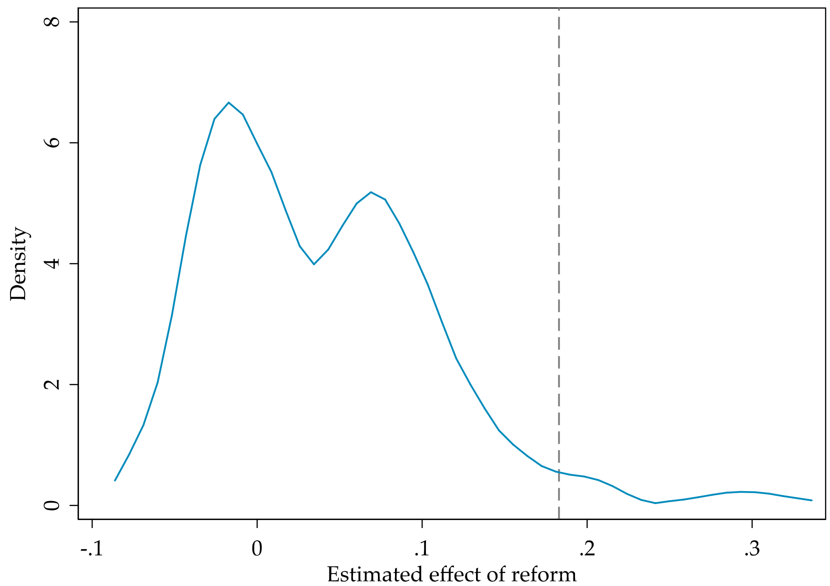
† The effect of reform is the linear combination of implementing reform and hosting a convention. In column (4), I subtract 12*monthly trend from the linear combination to remove the effect from the pre-existing trend.
Each regression includes asset and time fixed effects and an asset specific dummy that turns on at the end of the event window. Robust standard errors, clustered at the asset-level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 3 shows that when states are correctly grouped by default status the treatment effect falls in the 96th percentile. This suggests that randomly organizing states into two groups would not have produced the results in Table 3. Furthermore, organizing states by default status produces a treatment effect that is much larger than other logical groupings. For instance, states can be grouped by those that prevented the legislature from unilaterally increasing debt (53rd percentile), those that adopted a debt limit (60th percentile), those that required new debt to be accompanied by a tax increase (21st percentile), those that prohibited the lending of credit to private individuals or corporations (18th percentile), or those that adopted general incorporation

[29] I omit the monthly trend from resumption because states that did not default cannot be randomly included in the estimation of the resumption trend. Including the resumption trend but not removing 12*(trend coefficient) produces similar results.

laws (69th percentile).[30] In addition to producing a smaller treatment effect, none of these groupings produce an estimate that is statistically significant. This exercise further supports the claim that reform only conveys new information for those with a tarnished reputation by highlighting that only differences in default status explain whether a state was rewarded for adopting constitutional reform.

Figure 3: Distribution of treatment effects when default status is randomly assigned



Treatment effect is obtained using a variation of equation (1.1) where I include an indicator (interacted with "default" status) that is equal to one if the asset was issued by a state that hosted a constitutional convention by time t. The treatment effect is the linear combination of the convention indicator and the reform indicator, for "defaulting" states only.

Nine observations comprise the right tail of this distribution. These treatment effects are largely consistent with the hypothesis that only defaulting states were rewarded for adopting reforms. The "default" group in each of these specifications contains various subsets of the states that defaulted on their debts. Virginia is the only non-defaulting state to appear in these groupings, but it only appears in two of the nine specifications and it never appears by itself. Furthermore, for each of these nine specifications, there is no provision that is only adopted by

[30] The five states in the sample that adopted general incorporation laws are also the only states to prohibit the creation of corporations under special acts.

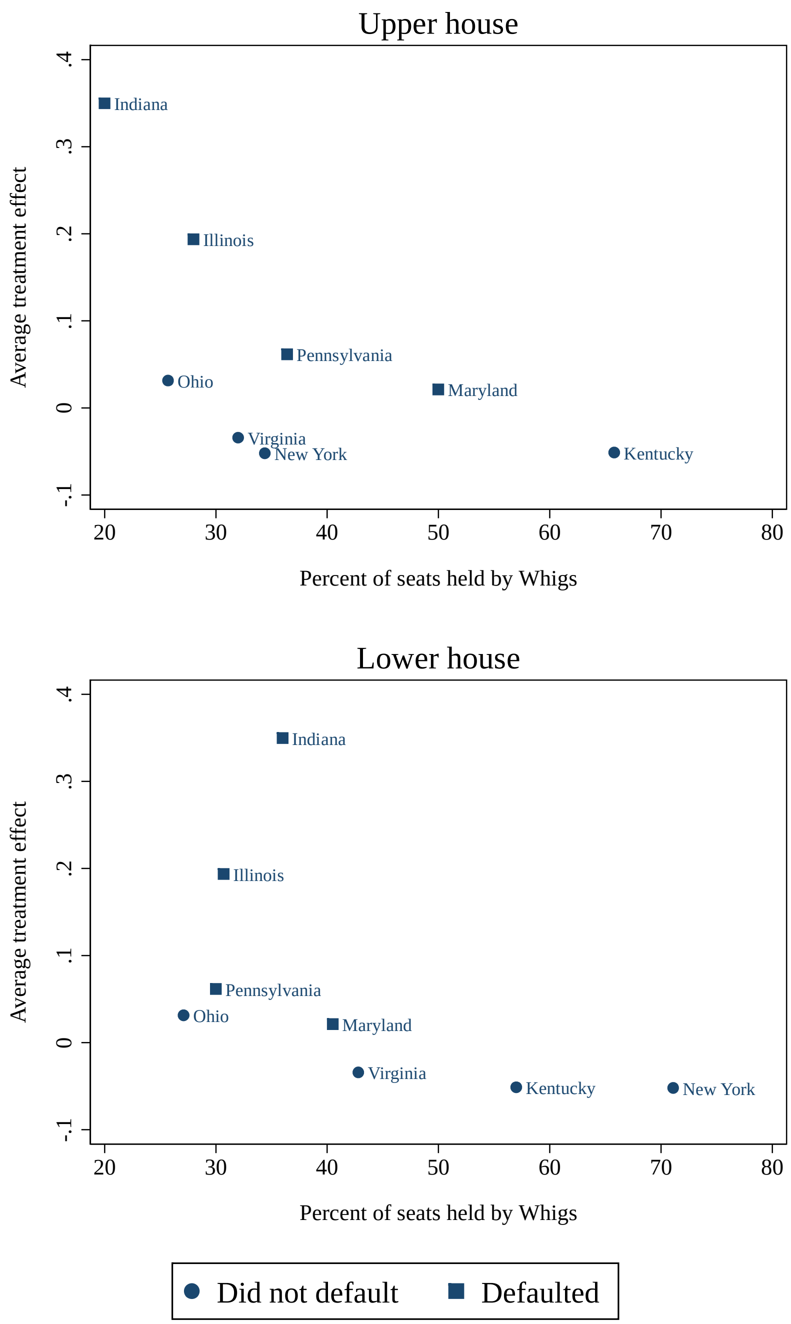
the “default” group. This suggests that the effect identified when states are grouped by default status cannot be explained by differences in the structure of the constitution.

1.4.c. Constitutional stability

The previous section showed that variation in the types of provisions adopted does not explain why only some states were rewarded. I did note, however, that defining the treatment group as a certain subsets of the defaulters resulted in a larger treatment effect. This highlights the possibility of heterogeneous treatment effects. In this section, I explore whether party politics might explain why some defaulting states were rewarded more than others. While I have implicitly assumed that states were equally committed to their reforms, this is not necessarily true. As Berkowitz and Clay (2005) note, states have had varied experiences with the stability of their constitutions. Louisiana, for instance, replaced its constitution in 1845 when Democrats controlled the legislature and again in 1851 when Whigs controlled the legislature. Louisiana’s experience, although extreme, suggests that party politics might have influenced the perceived stability of the adopted reforms.

In this section, I explore the role of constitutional stability by analyzing whether the share of Whigs within the legislature interacts with the market response. While the Democrats pushed for reform, the Whigs adamantly opposed it. This disagreement might have undermined the stability of the reforms. A low Whig share, for instance, might make it more difficult for the Whigs to gain control and amend, replace, or repeal the reforms. However, before exploring the role of party control, I must first estimate the average treatment effect (ATE) for each of the reforming states. This is achieved by estimating a variation of equation (1.1) where I use state-specific indicators for entering default, resuming payment, hosting a convention, and implementing reform. Each reforming state’s ATE is defined as the linear combination of the state specific coefficient for hosting a convention and implementing reform.

Figure 4: Share of Whigs within state legislature



Each state's treatment effect is obtained using a variation of equation (1.1) where I include an indicator that is equal to one if the asset was issued by a state that hosted a constitutional convention by time t. I interact the default indicator, resumption indicator, convention indicator, and reform indicator by state. The treatment effect is the linear combination of state specific coefficients from the convention indicator and the reform indicator.

In the first panel of Figure 4, I plot the average treatment effect for each state against the share of Whigs in the upper house of the state legislature.[31] There appears to be a negative relationship between the ATE and the share of Whigs within the state legislature, and consistent with the results presented thus far, this effect is driven by defaulting states. Figure 4 provides further evidence that financial markets did not respond to reforms adopted by non-defaulting states. This is illustrated by the fact that the treatment effects for non-defaulting states are clustered near zero. In the second panel of Figure 4, I plot the average treatment effect for each state against the share of Whigs in the lower house of the state legislature. Relative to the upper house, the relationship in the lower house is much less pronounced. This likely reflects institutional differences between the two houses (e.g. shorter term limits) that might undermine the durability of party control.

One might be concerned that defaulting states were rewarded for adopting constitutional reforms because those reforms interacted with some other institutional change. For example, suppose that following the debt crisis, states that defaulted were more likely to remove incumbents from office or more likely to elect Democrats. As a result of this change, when defaulting states adopted constitutional reforms they might have been rewarded, not because of their tarnished reputation, but because the constitution signaled the stability of the new government.

I address this concern in Table 6. Specifically, I estimate the following equation:

$$\begin{aligned} Whig_s(t) = \alpha + \beta_1 PostCrisis(t) + \beta_2 PostCrisis(t) \\ * \mathbf{1}[adopted\ reforms\ but\ did\ not\ default] + \beta_3 PostCrisis(t) \\ * \mathbf{1}[adopted\ reforms\ and\ defaulted] + [state\ FE's] + \varepsilon_s(t) \end{aligned} \tag{1.2}$$

where $Whig_s(t)$ is either the average Whig share in the upper house or lower house. Because it is unclear whether missing observations are coded as missing because there was not an election or because the data is genuinely missing, I focus on each state's average Whig share during two time periods – the six years preceding 1842 (the last year in which a state defaulted) or the six years following 1842. The variable $PostCrisis(t)$ is an indicator equal to one if the average was obtained for the post 1842 period.

The first column of Table 6 analyzes the share of Whigs in the upper house while the second column analyzes the share of Whigs in the lower house. In both specifications it appears

[31] Data on the share of Whigs was obtained from Burnham (1985).

that, in the years following the debt crisis, there was not a shift in party composition for any of the three groups; states that neither defaulted nor adopted constitutional reforms, states that did not default but did adopt constitutional reforms, or states that defaulted and adopted constitutional reforms. The results presented in Table 6 suggest that there was not a broader shift in party composition following the debt crisis.

Table 6: Party composition in state legislature

	Average share of seats in upper house controlled by Whigs	Average share of seats in lower house controlled by Whigs
Post crisis indicator for states that did not default or adopt reforms	-0.045	-0.015
	(0.030)	(0.023)
Post crisis indicator for defaulting states that adopted reforms	-0.089	-0.037
	(0.058)	(0.044)
Post crisis indicator for non-defaulting states that adopted reforms	0.024	0.012
	(0.051)	(0.038)
State fixed effects	Y	Y
R-squared	0.947	0.929
Observations	44	44

Standard errors in parentheses. Post crisis indicator is equal to one if year is greater than or equal to 1842. Sample restricted to the years 1834 through 1850, and the average is either for the years 1834-1841 or 1842-1850.
*** p<0.01, ** p<0.05, * p<0.1

The results presented in this section indicate that, consistent with my hypothesis, defaulting states were the only states rewarded for adopting constitutional reforms. These results also suggest that the political environment might affect the perceived stability of (and consequently the commitment to) the adopted reforms. Specifically, it appears that party control is correlated with the magnitude of the treatment effect. The states that benefited the most from reform were defaulting states where the share of Whigs was less than 40 percent. Because the two major parties of the time were the Whigs and the Democrats, as the share of Whigs fell below 40 percent the Democrats obtained a super majority. A Democrat super majority might have enhanced the credibility of the commitment by decreasing the likelihood that the Whigs would obtain enough seats to overturn the reforms. In other words, markets viewed a Democrat supermajority as a signal that the adopted reforms were stable.

1.5 REFORM AND BORROWING

Thus far I have shown that defaulting states were rewarded for implementing constitutional constraints but non-defaulting states were not. Specifically, I have shown that in the year following reform, assets for defaulting states increased by about 13 percent. This result cannot be explained by underlying trends or by differences in the types of provisions that states adopted. Furthermore, because price and the cost of capital are inversely related, this increase in prices indicates that, for defaulting states, borrowing costs fell following reform. In the next section I ask whether states benefited from lower interest rates by analyzing how reform affected total borrowing.

How might constitutional reforms affect total borrowing? Because these constitutional reforms were enacted in response to a debt crisis that resulted from states borrowing too much, one might expect that states that constrain themselves would borrow less in the future. On the other hand, the economy was largely back on track by 1845, and between railroad investments and borrowing for the Civil War, states likely found it desirable to maintain access to credit. Furthermore, the results in Section 1.3 indicate that the cost of borrowing declined for defaulting states that reformed their constitutions, which should also increase the quantity of debt demanded. When all of this is considered together, it appears that if capital markets reward states for adopting constitutional reforms, then one should expect to observe an increase in total borrowing.

1.5.a. Data and methodology

To understand whether reform affects access to capital, I gather data on total borrowing. Data on total outstanding debt in the years 1839, 1841, 1853, 1860, 1870, and 1880 are recorded in the "The Report on Valuation, Taxation, and Public Indebtedness". This data, which I transcribed from volume seven of the 1880 United States Census, will be used to explore whether reform affected total borrowing. Although observations occur at irregular intervals, the benefit of this dataset is that all states are represented. This dataset is also unique; data on outstanding state debts in the 19th century were often recorded at the state level in auditor or treasurer reports, which makes it difficult to construct a complete panel of state borrowing in the 19th century because state reports were not produced consistently.

As in Section 1.3, I use a differences-in-differences methodology to estimate how reform impacted total borrowing. Specifically, I estimate variations of the following equation:

$$\begin{aligned} Q_s(t) = \alpha + def(t) + res(t) + con(t) * \mathbf{1}[did\ not\ default] \\ + con(t) * \mathbf{1}[defaulted] + state\ FE's + year\ FE's + \varepsilon_s(t) \end{aligned} \tag{1.3}$$

where $Q_s(t)$, denotes the total outstanding debt per capita for state s in year t. The variables $def(t)$, $res(t)$, and $con(t)$ are indicator variables equal to one if state s entered default, resumed payment, or adopted constitutional reforms by time t. I interact $con(t)$ with default status to identify whether reputation influences the market's reaction.

Table 7: Constitutional reform's effect on total debt per capita

	(1)	(2)	(3)	(4)
Indicator for entering default	-39.314***	-42.836***	-39.739***	-41.414***
	(6.231)	(6.236)	(6.609)	(6.944)
Indicator for implementing constitutional reform *(tarnished reputation)*	15.587**	17.887***	18.567***	21.921***
	(6.140)	(6.202)	(6.198)	(6.766)
Indicator for implementing constitutional reform *(untarnished reputation)*	0.726	-1.302	-1.914	-1.406
	(2.964)	(2.939)	(2.962)	(4.107)
State fixed effects	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y
Geographic fixed effects post Civil War	N	Y	Y	Y
Years since default trend	N	N	Y	Y
Years since reform trend	N	N	N	Y

Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

1.5.b. Results

Table 7 presents the results from four variations of equation (1.3).[32] Column one, which estimates the specification presented in equation (1.3), indicates that total outstanding debt per capita increased by $15 for defaulting states that adopted constitutional reforms. But it is important to note that the act of default reduced a state's debt per capita by nearly $40. Nevertheless, these results imply that defaulting states were able to maintain better access to credit by adopting constitutional reforms. Specifically, default caused a $40 decline in per-capita borrowing and reform mitigated about 40 percent of that effect. As in Section 1.3, there appears to be no benefit for states that did not default.

The remaining columns in Table 7 introduce new controls but produce qualitatively similar results. In column two I add geographic fixed effects that turn on in 1870 to address the fact that Postbellum borrowing might have been influenced by the state's allegiance in the Civil War.[33] In column three I add a linear years since default trend, and in column four I include a linear years since reform trend. The coefficients for non-defaulting states are never significant, while the coefficients for defaulting states are always significant. These results reaffirm the findings in Section 1.3 – defaulting states benefited from reform while non-defaulting states were unaffected.

1.6 CONCLUSION

Can a sovereign improve its access to credit by adopting institutional constraints that reduce payment uncertainty? Previous empirical work has relied on case studies that analyze a single time series. America's 1840s state debt crisis, however, presents a unique opportunity to analyze this question with a panel of sovereigns. This setting, which allows me to control for general market trends, is better suited for inferring causality. By exploiting the plausibly exogenous timing in adoption of reform for eight American states, I find that defaulting states were

[32] It is important to note that, in contrast to Section 1.3, I now use the universe of states in my regression.
[33] I use the same geographic classifications presented in the 1880 Census where states were classified as New England, Middle, Western, or Southern.

rewarded with lower borrowing costs and increased access to capital following reform. Non-defaulting states, on the other hand, did not benefit from reform. These results cannot be explained by underlying trends or by differences in the content of the new constitution. There is some evidence, however, that party politics interacted with the market response for defaulting states. Historically, the Democratic Party controlled Indiana and Illinois – the two states that benefited the most from adopting reforms. Because the Whig Party opposed reform, Democrat control might have benefited defaulting states by signaling the stability of the new constitution.

Sovereign defaults occur regularly and at a great cost to both creditors and sovereigns (Tomz and Wright, 2013). The results presented in this paper suggest that sovereigns with tarnished reputations can benefit from establishing a commitment to debt repayment. Consistent with economic theory, I find that constitutional constraints are an effective mechanism for signaling commitment. These findings complements work by Stasavage (2008) and Mitchener and Weidenmier (2005 and 2010). In his analysis of Europe during the early modern period, Stasavage finds that interest rates were lower for sovereigns whose creditors wielded political power. These sovereigns established commitment by making it politically difficult to default. Mitchener and Weidenmier, on the other hand, study third party enforcement. They find that a credible threat of military intervention or economic sanctions is an effective way of enforcing payment and that markets value those threats. When considered together, it appears that both the formal constraints explored in this paper and informal constraints – constraints that raise the cost of undesirable behavior but do not prohibit it – are effective mechanisms for establishing commitment.

However, constraints are not the only mechanism for conveying commitment. Non-defaulting states demonstrated their commitment to debt repayment by incurring the costs necessary to avoid default. New York, for instance, avoided default by suspending its projects and reinstating its property tax. Ohio, which continued to finance its projects, increased its property tax rate by 240 percent between 1837 and 1845. Because non-defaulting states had already established their commitment, reform did not convey new information. Accordingly, these states were not rewarded for adopting reforms. Future researchers assessing the importance of institutional reforms will want to consider whether the reforms convey new information or if commitment had already been established through alternative mechanisms.

The results in this paper indicate that the fiscal provisions recommended in Title III of the European Fiscal Compact would likely benefit countries with tarnished reputations, at least with respect to the ability to borrow. However, there are specific long-run consequences that are not considered in this paper. For instance, how do these provisions affect a state's ability to act during a future economic crisis? Although balanced budget rules and debt constraints lower borrowing costs, if those constraints become binding during a downturn the state will be unable to implement fiscal policies that help mitigate the crisis. As another example, how might these constraints affect the provision of public goods? As Wallis (2000) argues, following 1842 the government investment became more decentralized, with local governments becoming responsible for investment in education, highways, water systems, and public utilities. Whether this arrangement was efficient is unclear. Analysis of these long-run consequences remains an important avenue for future research.

2.0 GRIDLOCK: ETHNIC DIVERSITY IN GOVERNMENT AND THE PROVISION OF PUBLIC GOODS

In this chapter, Daniel Jones and I ask how does diversity in government impact public good provision? To address this question, we construct a novel dataset linking the ethnicity of California city council candidates to election outcomes and expenditure decisions. We then study how the narrow election of a candidate whose ethnicity differs from the city's modal ethnicity affects expenditure decisions. We find that higher diversity on the council leads to less spending on public goods (especially in segregated cities) and fewer votes for affected council members when they run for reelection.

2.1 INTRODUCTION

Cities in the United States and elsewhere are becoming increasingly ethnically diverse. A growing body of empirical work has considered the political and economic consequences of this increase in diversity, highlighting the possibility for both positive and negative effects.[34] For instance, diversity has been shown to increase productivity, resulting in higher wages and residential values (Ottaviano and Peri, 2005, 2006). On the negative side, some have argued that diversity within a city may generate disagreement over the types of public goods government should provide. A consequence of this disagreement is a reduction in the amount the citizenry is willing to be taxed and a low level of government service provision. Empirical work on this front has led to mixed results.

[34] See Alesina and La Ferrara (2005) for a detailed survey of this literature.

Because it is ultimately a city's government that makes decisions on public spending, we analyze the extent to which diversity within government affects spending. This contrasts sharply with previous studies, which only consider diversity in the city as a whole. One might argue that the median voter theorem implies that it does not matter whether we explore the impact of the diversity of a city council or the diversity of a city itself. Specifically, the median voter theorem posits that vote-maximizing policymakers in a representative democracy (e.g., a city council) will disregard their own preferences and simply react to the wishes of the median voter. However, numerous theoretical and empirical papers have questioned the validity of the standard median voter model.[35] Consequently, how diversity in *government* impacts public good provision remains an important and unaddressed question.

The relationship between diversity in government and public spending is theoretically ambiguous. Within a governing body (in this case, the city council) diversity may indeed lead to "gridlock": disagreement over the type of public good to provide and a reduction in spending. However, there are also reasons to think that spending might increase. For instance, recent empirical work from a related literature documents that increased representation of a particular group can lead to more spending and transfers directed to that group (Pande, 2003; Cascio and Washington, 2014; Bhalotra, Clots-Figueras, Cassan, and Iyer, 2013). Thus, we might alternatively expect an *increase* in spending if diversity is accompanied by a greater push to fully serve each represented ethnic group. Some evidence consistent with this is provided by Rugh and Trounstine (2011). They find that diverse cities are more likely to propose spending packages that bundle public goods, which may be driven by an attempt to satisfy a diverse electorate. This can essentially be thought of as logrolling to satisfy distinct groups, which can of course lead to an inefficiently *high* quantity of public good (e.g., Weingast et al. (1981)). This theoretical ambiguity may help explain the mixed results in the existing literature, which has focused only on diversity at the city level.

To examine the relationship between diversity in government and public good spending, we study city councils in California. We construct a novel dataset that identifies the ethnicity of city council members and candidates in California from 2005 to 2011. This is then paired with

[35] See, in particular, Alesina (1988) for theoretical probing of the central implications of the median voter theorem, and Lee, Moretti, and Butler (2004) for an empirical assessment. Elsewhere, researchers have documented a variety of determinants of representative behavior beyond simple response to the preferences of the median voter (Levitt, 1996; Washington, 2008).

detailed annual city budgets, obtained from the California State Controller's Office. One advantage to studying within-government diversity is that city council elections naturally lend themselves to a quasi-experimental design.[36] Because council members are elected, we focus on close elections that could potentially shift the diversity of the council (i.e., an election between members of two different ethnic groups). In close elections, the winner is plausibly random and, as a result, so too is the resulting change in diversity. We show that the narrow election of a candidate whose ethnicity is *not* the city's modal ethnicity is a plausibly exogenous shock to diversity within the city council. The narrow victory is associated with increased diversity within the city council but not at the city level. Relying on this fact, we implement a regression discontinuity design that allows us to measure the extent to which these random shocks to diversity affect spending on public goods. Our results indicate that diversity leads to gridlock; cities reduce the amount they spend on public goods as their city council becomes increasingly diverse. We also find that members of a council that experienced an exogenous shock to diversity receive fewer votes when they run for re-election. This latter point suggests that the city's population is dissatisfied with the decline in public goods, ruling out the possibility that diverse councils simply achieve greater efficiency in public good provision.

These results speak to the existing literature on diversity within a city and public good provision. This literature was essentially started by Alesina, Baqir, and Easterly (1999), who present the theoretical justification for a decline in public good spending, which was noted above. Their model assumes that different ethnicities have different preferences over public goods and that decisions about public goods are made in a two-stage election process. In the first stage, the type of public good is determined. In the next stage, the amount of funding for the public good (and therefore the amount citizens will be taxed) is decided. In a median voter framework, greater heterogeneity leads to the adoption of a "compromise" public good. In the next stage of voting, the median voter is not willing to pay high taxes on a public good that is distant from their preferred type. In a homogeneous city, everyone agrees on the type of public good being offered, and the median voter is willing to pay a high tax bill to fund it.

[36] This contrasts with previous literature, which has struggled to find a source of quasi-random variation in the diversity of the city. An exception is Dahlberg, Edmark, and Lundqvist (2012) who take advantage of random assignment of international refugees to localities in Sweden, but they address a different question.

Analyzing a cross section of U.S. cities, Alesina et al. document a (mostly) negative relationship between diversity and public good spending.[37] Recent empirical work has questioned the robustness of this relationship in a panel (Boustan et al., 2013; Hopkins, 2011) or with additional controls (Gisselquist, 2013). These studies have found no clear and consistent connection between diversity and public good spending.[38] By taking a different angle and exploring the way that diversity in *government* impacts government spending decisions, we hope to bridge this gap in the literature. Ultimately, our results suggest that even in the context of a small decision-making body, where one might expect favor-trading and an increase in spending, Alesina et al.'s predictions are realized.

Our results also speak to a more general literature on the impact of diversity within small groups or organizations, which has been studied in economics, psychology, and human resources. That literature tests competing hypotheses: diversity may lead to disagreement and a decline in performance, or diversity may lead to a variety of skills and ideas, generating improved performance. Results are mixed. In a randomized laboratory experiment, McLeod et al. (1996) find that diverse groups reach solutions to brainstorming tasks that are judged to be of higher quality. Shore et al. (2009) review work mostly from Human Resources researchers. There, the disagreement hypothesis seems to dominate; most of the reviewed research points to a negative relationship between diversity and performance. La Ferrara (2005) studies production cooperatives in Africa. She finds that ethnic diversity leads to *lower* likelihood of workers specializing in different tasks, perhaps also pointing to an inability to agree and coordinate on the most efficient method of production. Ben-Ner et al. (2014) find some evidence of both effects, potentially reconciling mixed results. Drawing on theories from psychology and elsewhere, they test the idea that diversity can be positive when an organization is threatened and negative when the organizational goal is self-promotion. They test these theories by comparing diversity of soccer players on offense and defense, finding that diversity has a positive impact for defensive players (who are under threat of opposing players' goal scoring) and a negative impact for offensive players (whose "organizational goal" is promotion of the team through goal-scoring.) In this light, it is worth keeping in mind that the production of new public goods may be viewed

[37] They find that diverse cities spend less on productive goods (education, roads, and sewerage) but spend more on police protection. Some of their specifications attempt to account for the potential endogeneity of diversity using lagged diversity as an instrument, which yields similar results.

[38] It should be noted that Boustan et al. (2013) revisit the relationship between diversity and public good expenditures as a subsection of a paper otherwise addressing a different question.

as a “promotion” task, and that we may be neglecting outcomes wherein diversity of the council has the potential to be positive.

2.2 CITY COUNCILS IN CALIFORNIA

California state law provides a number of guidelines for the institutional structure of municipal governments. Specifically, city councils must contain five councilmembers, and councilmembers are to be elected “at-large” during a general municipal election. Councilmembers serve staggered four-year terms, with elections every two years. Elections are nonpartisan, so political party is neither observed to the voters nor is it in our data. California state law defines the mayor as simply another member of the city council and does not provide for any additional powers. In these “council-manager” cities, the council (including the mayor) dictates policy for the city, which is in turn carried out by the city manager. Furthermore, the mayor is typically selected by the city council from amongst its own members.

There are two ways that a city can deviate from the above guidelines. If the city is “general law” – the default form of government for incorporated cities – then it can submit a ballot measure to be approved by the electorate. For “chartered” cities, any deviation must be specified in the city’s charter.[39] Nevertheless, a 2006 survey conducted by the International City/County Management Association (ICMA) provides a number of statistics illustrating that most cities conform to the state’s guidelines.[40] Specifically, 93 percent of cities are council-manager cities and the mayor serves on the city council for 98 percent of the cities – because of

[39] Chartered cities differ from “general law” cities in one important aspect – chartered cities have supreme authority over all municipal affairs while “general law” cities are bound by the state’s general law. The process for adopting, amending, replacing, or repealing a charter is quite involved. Specifically, conditional on receiving signatures from at least 15 percent of registered voters within the city, a city can hold an election that asks voters whether a charter commission should be elected to propose a new charter and which candidates should serve on that commission. Assuming the charter commission is elected, the charter prepared by the commission must then be approved/ratified by voters in the next election. California state law also dictates that the ballot description must enumerate new city powers resulting from the adoption of the new charter. Alternatively, the governing body can motion to propose, amend, replace, or repeal a charter. California state law imposes several requirements aimed at informing the public of the proposed changes, but aside from those provisions, the process for accepting/ratifying those changes is exactly the same.

[40] The survey in question is ICMA’s 2006 *Municipal Form of Government* survey, which is the source of all of the statistics in this section.

this, when we calculate the ethnicity of the city council we include mayors. 88 percent of cities only have five councilmembers and councilmembers are elected at-large for 92 percent of cities.

A city's institutional structure tends to be relatively stable over time. For instance, in the five years preceding the survey fewer than seven percent of cities attempted to alter their form of government. When cities do attempt to alter their form of government, it is typically to switch from at-large to district based elections (or vice versa). Many of these attempts, however, were ultimately unsuccessful.

2.3 DATA

In this paper, we study the link between city government spending and ethnic diversity within city government. We rely on three broad sources of data to identify: (1) the names and vote totals of individuals who served on a city council or ran for city council but lost; (2) the ethnicities of those councilmembers and candidates; and (3) how much the city council spent and the allocation of those expenditures amongst various categories.

The first source of data (and the reason we focus on California) is the *California Election Data Archive* (CEDA), which provides the names and number of votes for every candidate in every local government election occurring between 1995 and 2011. Because ethnicity is not listed in this dataset, we supplement the CEDA election returns with novel data. Specifically, we collect pictures of councilmembers and candidates from candidate websites, newspaper articles, and other sources; we then conduct a survey on Amazon's "Mechanical Turk" website where we ask workers to report the candidate's ethnicity based on the candidate's name and picture. Together, these datasets allow us to identify the ethnic composition of city councils and the counterfactual composition (what the composition would have been had the losing candidate won). Finally, we obtain expenditure data from the California State Controller's Office city budget records. These records provide detailed annual accounts of expenditures and revenues for every city in California between 2005 and 2011.

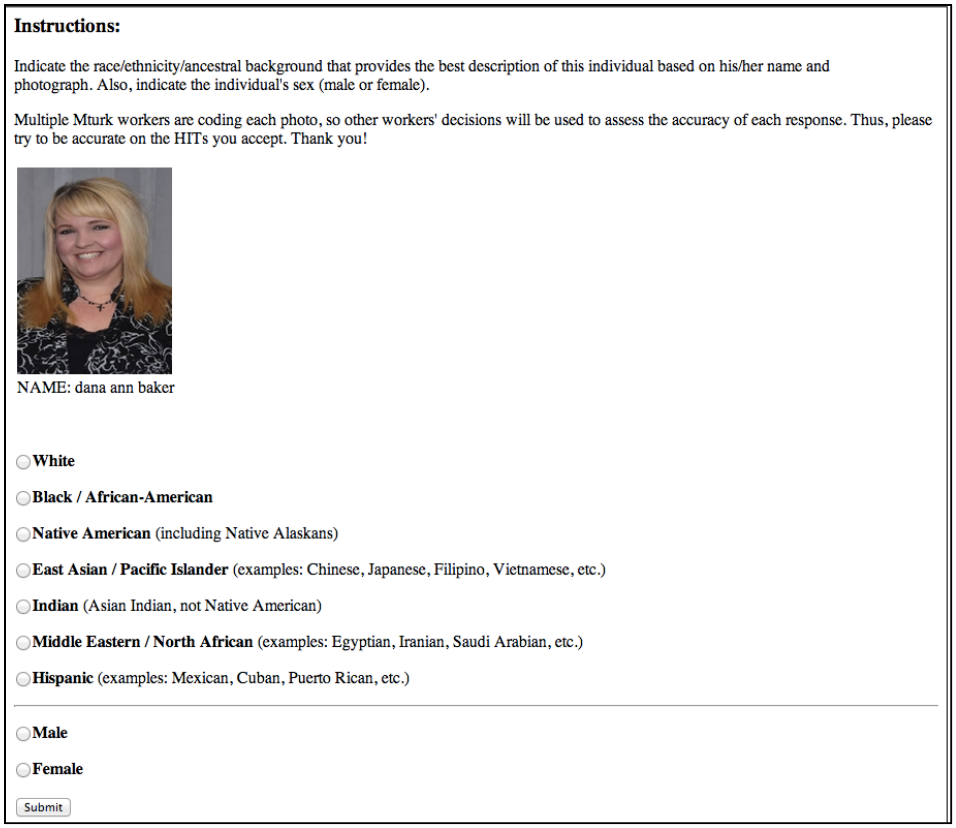
2.3.a Ethnicities

Because of the importance and novelty of ethnicity-identification in our data, this subsection describes our procedure for collecting these data. We focus our data collection efforts on the intersection of the two California datasets. That is, we only locate photos for those that served on the city council between 2005 and 2011 and those that ran for city council but just lost. We successfully located photos for 3,966 candidates. After collecting these photos we conduct a survey on Amazon's "Mechanical Turk" website where we ask workers to report the candidate's ethnicity based on the candidate's name and picture.[41] The worker can choose from the following options: White, Black, Native American, East Asian or Pacific Islander, Indian, Middle Eastern or North African, or Hispanic. We also ask the worker to identify the gender of the candidate. Figure 5 provides a screenshot of the task.

We collect ten unique responses for each candidate. There was no limit to the number of photographs a worker could code, but they never observed a candidate more than once. For the sake of incentive compatibility, workers were told that the responses from other workers would be used to judge the accuracy of their work. Specifically, workers were not paid unless a majority of their responses matched the modal response. Furthermore, because the task is relatively straightforward, the marginal cost from answering truthfully, relative to randomly selecting ethnicity and gender, was small.

[41] Specifically, we ask the worker to indicate the race/ethnicity/ancestral background that provides the best description of the individual based on their name and photograph.

Figure 5: An example task



Instructions:

Indicate the race/ethnicity/ancestral background that provides the best description of this individual based on his/her name and photograph. Also, indicate the individual's sex (male or female).

Multiple Mturk workers are coding each photo, so other workers' decisions will be used to assess the accuracy of each response. Thus, please try to be accurate on the HITs you accept. Thank you!

NAME: dana ann baker

- **White**
- **Black / African-American**
- **Native American** (including Native Alaskans)
- **East Asian / Pacific Islander** (examples: Chinese, Japanese, Filipino, Vietnamese, etc.)
- **Indian** (Asian Indian, not Native American)
- **Middle Eastern / North African** (examples: Egyptian, Iranian, Saudi Arabian, etc.)
- **Hispanic** (examples: Mexican, Cuban, Puerto Rican, etc.)

- **Male**
- **Female**

Submit

We use the modal response as the ethnicity of the candidate if at least five of the ten workers agreed with the modal response. This restriction leaves us with 3,944 identified candidates. The average rate of agreement for these candidates was 94 percent, which implies that on average 9.4 of the ten workers chose the same ethnicity. The average rate of agreement for gender was 99 percent. To assess the accuracy of these responses, we solicited ethnicity information from city clerks and a number of organizations concerned with minority representation. We contacted 450 cities, and of the 230 cities that responded, 96 provided information regarding candidate ethnicity. Specifically, we received the ethnicity for 1,194 councilmembers. We also received responses from the *National Association of Latino Elected and Appointed Officials* and the *Asian Pacific American Institute for Congressional Studies*. Together, these organizations provided us with the ethnicity of 361 Asian and Hispanic candidates.

There is a considerable amount of overlap between these three sources (Amazon's "Mechanical Turk", city responses, and responses from ethnic organizations). Table 8 summarizes the extent to which these sources agree with each other. 271 of the 3,944 ethnicities obtained from the pictures we collected were also listed in the information provided from cities, and 320 were listed in the information obtained from ethnic organizations. The responses provided by Amazon's Mechanical Turk matched the information provided by cities 94.83 percent of the time, and the information provided by cities matched the information provided by ethnic organizations 91.8 percent of the time. Our lowest match rate, 77.19 percent, comes from comparing the overlap between the Mechanical Turk responses and the information provided by ethnic organizations. These mismatches occur when the Mechanical Turk chooses "White" instead of "Asian" or "Hispanic", highlighting the importance of obtaining ethnicity information for councilmembers and candidates whose ethnic background might be mistaken for White. Consequently, we use the information provided by cities as the true ethnicity whenever possible. We then rely on the lists obtained from ethnic organizations to identify any remaining Hispanic or Asian candidates. This further increases the accuracy of the Mechanical Turk responses because we are only relying on their ability to determine whether a candidate is White, Black, Native American, Indian, or Middle Eastern.

Table 8: Source overlap

	Count	Match percentage
Mechanical Turk and city	271	94.83%
City and list	122	91.80%
Mechanical Turk and list	320	77.19%

See text for an in depth discussion of each source

Our final sample includes ethnicities for 4,161 of the 4,788 councilmembers and candidates that either served on the city council between 2005 and 2011 or ran for city council but just lost. Put another way, our sample allows us to identify the ethnicity of each councilmember for 1,774 of the 2,321 council-year pairs between 2005 and 2011.

2.3.b Measuring diversity

To measure diversity we focus on fractionalization and polarization, the most prominent measures used within the literature. Fractionalization is calculated using the following equation:

$$Fractionalization_{ct} = 1 - \sum_{e} (share_{cte})^2 \qquad (2.1)$$

where $share_{cte}$ is the share of the city council in city *c* during year *t* that is of ethnicity *e*. The fractionalization index ranges from zero to one, where zero implies that all city councilmembers are of the same ethnicity. The polarization index, as proposed by Reynal-Querol (2002), is designed to take into account the potential for group conflict by measuring the extent to which groups fall into two distinct categories. The polarization index is measured using the following equation:

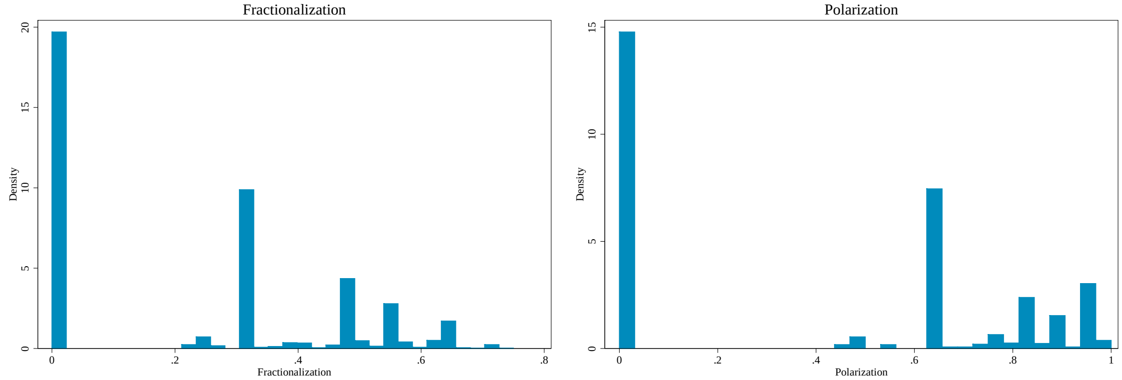
$$Polarization_{ct} = 4 * \sum_{e} (share_{cte})^2 \, (1 - share_{cte}) \qquad (2.2)$$

where, as in Equation (2.1), $share_{cte}$ is the share of the city council in city *c* during year *t* that is of ethnicity *e*. Similar to fractionalization, the polarization index ranges from zero to one, where zero implies that all councilmembers are of the same ethnicity.

The difference between fractionalization and polarization is best illustrated by the following thought experiment. Suppose that there are five councilmembers of different ethnicities. If the council added a sixth member from yet another ethnic group, fractionalization would increase while polarization would decrease. This is because fractionalization is maximized when each councilmember is of a different ethnicity. Polarization, on the other hand, is maximized when the seats are distributed into two ethnic groups, leaving the most potential for conflict.

Figure 6 plots the distribution of both indices for our sample, and Table 9 summarizes the ethnic composition of those councils. The typical council has five seats, where 3.9 of those seats are held by white councilmembers. Fractionalization ranges from 0 to 0.75. The mean is 0.229 and the median is 0.32. Polarization ranges from 0 to 1 with a mean of 0.408. The median is 0.64.

Figure 6: Distribution of diversity within city council



Polarization and fractionalization were only computed for council-year pairs in which ethnicity information for all councilmembers was obtained. Sample includes 1,774 of the 2,321 council-year pairs appearing between 2005 and 2011.

Table 9: Composition of city councils

	Mean	Median	Min	Max	Standard deviation	Number of observations
Fractionalization	0.229	0.32	0	0.75	0.231	1774
Polarization	0.408	0.64	0	1	0.392	1774
Number of seats	5.300	5	4	11	0.982	1774
White councilmembers	3.907	4	0	9	1.653	1774
Hispanic councilmembers	0.950	0	0	7	1.428	1774
Black councilmembers	0.203	0	0	5	0.578	1774
Asian councilmembers	0.198	0	0	4	0.544	1774
Middle Eastern councilmembers	0.021	0	0	3	0.161	1774
Indian councilmembers	0.019	0	0	2	0.147	1774
Native American councilmembers	0.003	0	0	1	0.053	1774

See text for discussion of how councilmember ethnicities were obtained.

2.3.c Outcome variables

Our outcome variables are drawn from annual city budgets, which we obtained from the California City Controller's Office. These datasets report detailed expenditure and revenue categories for every city in California. For instance, we observe the amount spent on parks, servicing debts, police, etc. in addition to revenue from various sources (general revenue, intergovernmental revenue, etc.).

Ultimately, we are interested in the impact of diversity on public good spending. One challenge is that there is substantial variation across cities in the types of public goods offered. To achieve uniform measures of spending across cities, we collapse spending into three broad categories: government administration, debt repayment, and public goods. The government administration and debt repayment categories are provided by the Controller's Office. "Government administration" includes: legislative expenditures, management and support, and all salaries and benefits to workers within either of these categories. "Public goods" is a category of spending we created by taking a city's total expenditures for the year and removing expenditures on "government administration" and debt repayment. The "Public goods" category therefore includes all spending on roads, parks, police protection, sewerage, public transportation, etc. This aggregated measure of public good expenditures provides us with a common measure across all of the cities in our sample.

For the sake of comparison, Table 10 reports summary statistics for all city-year pairs reported in the California Controller dataset and for the city-year pairs where we were able to obtain ethnicity information for all of the city councilmembers. Although the cities with completed councils are slightly larger, ethnic composition and spending patterns across the two samples are similar. Both samples are roughly 50 percent white with city-level fractionalization of about 0.5 and city-level polarization of 0.70. Furthermore, both samples allocate approximately 86 percent of their revenues to public goods, 13 percent to government administration, and 1 percent to debt repayment.

Table 10: Summary statistics

	All cities	Completed councils
Population characteristics		
Total population	63307.890	76488.750
White share	0.485	0.462
Hispanic share	0.341	0.348
Asian share	0.103	0.118
Black share	0.037	0.040
Native American share	0.006	0.004
Other share	0.002	0.002
Fractionalization within city	0.490	0.502
Polarization within city	0.702	0.709
Government spending		
Per-capita public good expenditures	\$5,845.43	\$5,023.72
Per-capita debt expenditures	\$1,404.42	\$1,628.29
Per-capita gov. admin. expenditures	\$553.98	\$488.16

Sample of cities includes each of the 2,321 city-year pairs that reported expenditure information to the California City Controller's office between 2005 and 2011. Completed councils restricts to the 1,774 council-year pairs where the ethnicity is known for each councilmember. Population characteristics are obtained by interpolating between the 2000 and 2010 decennial Censuses. See text for description of government spending categories.

2.4 EMPIRICAL APPROACH

2.4.a Empirical approach and data

To assess the relationship between diversity and public good spending in our panel of data, we might simply regress spending on fractionalization or polarization (and include city fixed effects). Of course, in doing so, we may be concerned about endogeneity between diversity and spending. For instance, perhaps a city is experiencing a period of growth, with a more diverse population moving in and electing a diverse government. If this is the case, diversity might increase at the same time that spending on public goods increases to accommodate an expanded population, but the increased diversity *within* government does not necessarily *cause* the increased spending.

To deal with potential endogeneity, we use close elections with the potential to impact council-level diversity as a source of random variation.[42] Specifically, we implement a regression discontinuity design, focusing on elections between a candidate whose ethnicity differs from the modal ethnicity of the city and a candidate whose ethnicity matches the modal ethnicity. We do so under the assumption that the election of the "non-modal" candidate increases the diversity of the council. This assumption is tested and overwhelmingly confirmed in the next section.[43]

We restrict our sample to cities that experienced an election between a "modal" and a "non-modal" candidate *at any point* between 2006 and 2009, as these are the cities where an election has the potential to change the diversity of the council.[44] For these cities, we construct a panel that spans from fiscal year 2005-06 until fiscal year 2010-11.[45] A city faces the potential for becoming "treated" following the first election that meets these criteria. Prior to this, all cities are considered "untreated". If the non-modal candidate wins the election, then the city is treated; in the notation of our empirical models, an indicator called "non-modal wins" is set to 1. If the modal candidate wins, the city remains untreated and "non-modal wins" remains 0. Treatment begins in the fiscal year *following* the election, as it is only then that the candidate has the opportunity to impact the budget. For example, the first budget that a candidate elected in November of 2006 will have any input on is the 2007-2008 fiscal year budget. Thus our non-modal wins indicator would remain zero for fiscal year 2006-2007.

It is important to note that our outcome variable of interest (within-government diversity) is not binary – the extent to which any candidate affects the diversity of the city council depends both on the candidate's ethnicity and the ethnicity of the other councilmembers. Although using "non-modal win" as an instrument for within-government diversity would better accommodate this fact, we focus on the binary treatment variable non-modal wins because it allows us to closely match the approach taken by Ferreira and Gyourko (2009, 2014) in their evaluations of

[42] The idea of using close elections as a source of random variation in political composition was made famous by Lee et al. (2004). Recent work analyzing elections in the U.S. House of Representatives (Caughey and Sekhon, 2011; Grimmer et al., 2011) has challenged the validity of this design, arguing that even close elections are often nonrandom. However, Eggers et al. (2014) cast doubt on this claim. Eggers et al. study elections at several levels and across time periods, and find no evidence that incumbents are more likely to win in a close election.

[43] We focus on modal ethnicity because there are many cities without a clear majority group. If a member of a non-modal ethnic group wins when facing off against a member of the modal group, we expect the diversity of the council to increase.

[44] For cities that hold district-based elections, each seat is decided by a separate election. Because our empirical strategy requires at most one election for each city-year pair, we use the closest election between two candidates of different ethnicities as the election of interest.

[45] The fiscal year runs from July to July.

the impact of a Democrat or female mayor being elected. Nevertheless, in DIG_section_5.3 we implement an instrumental variables approach and find similar results.

Before proceeding to our estimating equation, it is important to talk about an issue related to cities that experience more than one potentially-treatment inducing election. To take advantage of the panel structure of the data, we would like to see government spending before and after an opportunity for treatment. Note, however, that in our panel setup, a city not yet "treated" should *not* have a margin of victory of zero. Although coding the margin as zero before the election may seem natural, doing so would imply that the counterfactual to a narrow non-modal win is not just a non-modal loss (which is our intention), but also all observations prior to the election. In other words, a *city* is defined throughout the panel as being a "close election" city, but the *impact* of this election is identified only after it is happened.

Of course, setting the "margin of victory" as constant throughout time generates complications if a city experiences more than one election that might cause a shift in diversity. Thus, for some cities, it is necessary to truncate their panel so that each city in our sample only has the potential to be treated once. For the 40 cities in our sample that experience a second election between a modal and non-modal candidate, we truncate their panel, dropping all observations coinciding with and following the year that the second of the potentially treatment-inducing election occurs.

2.4.b Empirical model

For our analysis, we will estimate variations of the following equation:

$$\begin{aligned} y_{ct} \quad = \quad & \propto + \beta_1 \mathbf{1}[Non-modal\ wins_{ct}] + \beta_2 F(margin\ of\ victory_c) \\ & + \beta_3 \mathbf{1}[Non-modal\ wins_{ct}] * F(margin\ of\ victory_c) \\ & + year\ FE's + city\ FE's + \varepsilon \end{aligned} \qquad (2.3)$$

where the subscript "c" indicates the city and subscript "t" indicates the year. A candidate is classified as "non-modal" if their ethnicity does not match the modal ethnicity within the city.[46] The variables $\mathbf{1}[Non-modal\ wins_{ct}]$, $F(margin\ of\ victory_c)$, and the interaction of the two mirror the relatively standard parametric regression discontinuity approach. Also note that, as is relatively standard in this approach, margin of victory enters through *F(.)*, which is a polynomial

[46] The modal ethnicity within a city is drawn from decennial Census data. Because we do not have yearly data for this variable, we use the modal ethnicity from the midpoint of our sample, 2008. The modal ethnicity in 2008 was calculated by interpolating ethnic shares between the 2000 and 2010 censuses.

function. When we report results, $F(.)$ will either be a second, third, or fourth-degree polynomial. The dependent variable, y_{ct}, will capture either: per-capita public good spending, per-capita debt servicing, or per-capita spending on government administration. Standard errors are clustered at the city council level.

"Margin of victory" is simply the difference between the vote share received by the winner and the vote share received by the loser.[47] Because of this, β_1 can be interpreted as the impact of a non-modal victory when the margin is zero; in practice, we interpret this as the impact of a victory in a very close election. Thus, the "non-modal wins" coefficient is of primary interest, which can be interpreted as the causal impact of a non-modal victory. Of course, the time invariant nature of "margin of victory" implies that it is ultimately absorbed by the city fixed effects and the coefficient β_2 in Equation (2.3) will be omitted.

2.5 EMPIRICAL RESULTS

2.5.a Assessing non-modal victory as an exogenous shock to diversity

In this section, we conduct tests to ensure that our research design generates an exogenous shock to diversity within government. Our first test illustrates that the election of a non-modal candidate increases diversity within the city council. The second test analyzes the relationship between our forcing variable (non-modal win margin) and other outcome variables to ensure that a narrowly elected non-modal candidate is indeed an exogenous shock. To assess both issues, we estimate specifications following the structure of Equation (2.3), but taking measures of council-level diversity and measures of both city-level diversity and partisanship as the outcome variables. Essentially, we aim to show that there is a discontinuity in council-level diversity at the cutoff, and there is not a discontinuity in other observables.

Table 11a can be thought of as a "first stage," documenting that a non-modal victory indeed shocks the diversity of the council. Regardless of whether we measure diversity with

[47] In elections with "multiple winners" (as in a city council election to fill multiple seats), the margin victory is measured for marginal candidates: the difference between vote share of the last-placed winner and the first-placed loser.

fractionalization (column one) or polarization (column two), there is a strong and positive relationship between the election of a non-modal candidate and the diversity of the city council. To get a sense of the magnitude of the coefficient, consider shifting from a council with four White members and one Hispanic member to a council with three White members and two Hispanic councilmembers. This would change fractionalization from 0.32 to 0.48, which is roughly equivalent to the magnitude observed for the "non-modal" indicator in Table 11a.

Table 11a: The relationship between a narrow non-modal victory and council diversity

	(1) Fract.	(2) Polar.
Non-modal wins	0.104**	0.180**
	(0.042)	(0.073)
City fixed effects	Y	Y
Year fixed effect	Y	Y
Observations	503	503
R-squared	0.866	0.844

Robust standard errors (clustered at council-level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 11b can be thought of as a test of an exclusion restriction assumption (if we were using an instrumental-variable approach). Our general goal is to test the impact of council-level diversity on public spending. An increase in council diversity that is simply driven by, for instance, an upward trend in population-level diversity may have its own impact on public spending.[48] But if our "non-modal wins" indicator merely reflects changes in population-level diversity, then our methodology does not represent a random shock to diversity of the decision-makers, and therefore does not represent a methodological improvement over the existing literature. Thus, we regress "non-modal wins" on city-level fractionalization in column one and city-level polarization in column two. Note that city-level fractionalization and polarization are constructed from decennial census data, which means they are interpolated for intercensal years.

[48] In fact, this is precisely the mechanism that is implicitly assumed by Alesina et al. (1999) that allows them to analyze city-level diversity instead of within government diversity.

However, these interpolated measures still allow us to pick up population-level trends. We find that a “non-modal win” is not related to these population-level diversity measures.

In the remaining columns of Table 11b, we assess whether there are systematic differences in political composition of the city to the left and right of the cutoff. For instance, if “non-modal” candidates are associated with a particular party (even though city council elections are non-partisan), we might expect those candidates to get a boost in close elections in cities where that party is strong[49], which may then suggest our results are driven not by a shock to diversity but by a shock to partisan affiliation. To address this, we rely on voter registration records to assemble an annual city-level panel of voter registration (and party affiliation) in California.[50]

Using these data, we find that the following observables are smooth around the cutoff: share of registered voters who are registered as Democrats (column 3), share of registered voters who are registered as Republican (column 4), share of citizens who are registered voters (column 5), and a measure of political competition (column 6). The “population share of registered voters” is meant to provide a proxy for civic engagement. The measure of political competition is taken from Besley et al. (2010) and reaches a maximum when there is exactly the same number of Democrats and Republicans in a city.

[49] More generally, we suspect that it is unlikely that our results are confounded by unobserved correlation between “non-modal” status and affiliation with a particular political party. Because modal ethnicity is city specific, even if there is a correlation between particular ethnicity and a particular political party, that ethnicity is likely to be code as “modal” in some cases and “non-modal” in others.

[50] The California Secretary of State website provides, for even-numbered years, the number of registered voters in the city, as well as the number of voters registered as Republican, Democrat, or some other party. We constructed an annual panel by linearly interpolating these totals.

Table 11b: The relationship between a narrow non-modal victory and city-level observables

	(1)	(2)	(3)	(4)	(5)	(6)
	City fract.	City polar.	City Democrat share	City Republican share	Pop. share registered voters	Political competition
Non-modal wins	0.000	-0.003	0.002	-0.001	-0.000	-0.004
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.005)
City fixed effects	Y	Y	Y	Y	Y	Y
Year fixed effect	Y	Y	Y	Y	Y	Y
Observations	503	503	559	559	559	559
R-squared	0.866	0.844	0.992	0.995	0.990	0.981

Robust standard errors (clustered at council-level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

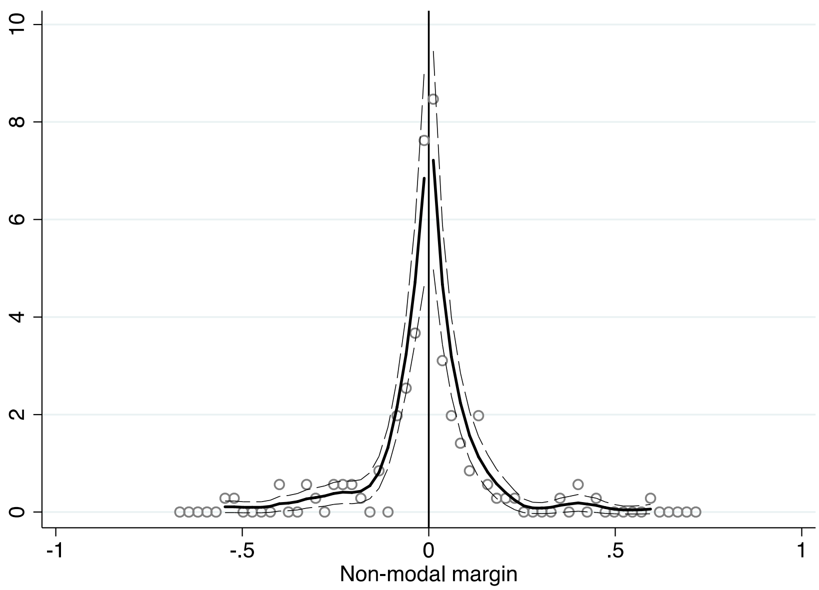
Next, we address a common concern in regression discontinuity designs: the "forcing variable" (in this case, non-modal margin of victory) should be balanced around the cutoff (the point where one candidate barely wins). If we define the "forcing variable" as [*non-modal margin=non-modal vote share – modal vote share*], this implies that there should be roughly the same number of observations to the left of *non-modal margin* as there are just to the right of *non-modal margin*.

This issue is especially important to our research design. Not only have some questioned the "randomness" near the cutoff when implementing regression discontinuity designs to electoral outcomes (Caughey and Sekhon, 2011; Grimmer et al., 2011), but Vogl (2014) documents concerns specifically in the context of race and city politics. He finds that in the South, but not the North, there is clear evidence that black mayoral candidates are slightly more likely than white candidates to win a close election.

Here, we have a roughly equal number of observations on either side of the cutoff. This is first documented graphically in Figure 7. In Figure 7, we follow McCrary (2008) and plot a discontinuous density function around the cutoff (non-modal margin=0). Figure 7 is similar to graphs used by Vogl (2014).[51] The figure demonstrates that the density just to the left of the cutoff is statistically indistinguishable from the density just to the right of the cutoff.

We can also document that modal are not more likely to win close elections using simple statistical tests. Ideally, when the election is close, the probability of a non-modal victory should be 0.5. When an election is decided by a margin of 5 percent of less, the observed proportion of non-modal victories is 0.5062. Using a binomial test, this is statistically indistinguishable from 0.5 (the p-value is effectively one).[52]

Figure 7: Distribution of non-modal margin of victory



The x-axis represents the "Non-modal margin" of victory (non-modal vote share – modal vote-share). The y-axis represents the density. Solid lines are estimates, while dashed lines are standard errors.

Thus, we feel confident that our design yields a shock that (a) is strongly correlated with council-level diversity and (b) is indeed randomly determined. We now proceed to our estimates of the relationship between diversity and government spending.

[51] In fact, our figure is constructed using code from McCrary's 2008 paper, which is available on his website.

[52] The same holds when we tighten the definition of a "close election". When an election is decided by a margin of 2 percent of less, the observed proportion of non-modal victories is 0.4898. Using a binomial test, this too is statistically indistinguishable from 0.5.

2.5.b Main results

Prior to reporting the results from our main regression-discontinuity estimations, it is perhaps informative to start with a simpler fixed-effects regression that does not attempt to deal with endogeneity between diversity and public good spending. These results are reported in Table 12. As seen in Columns 1 and 4, there is no statistical relationship between diversity and per-capita public goods expenditures.[53] This result is reassuring, as these specifications are close to those of Boustan et al. (2013) and Hopkins (2011) in that they assess the relationship between diversity and public good spending in a panel. As in those papers, there is no clear systematic relationship between the two variables. Similarly, there are no shifts in the amount spent on debt management of government administration.

Table 12: Correlational relationship between council-level diversity and government spending per capita

	(1)	(2)	(3)	(4)	(5)	(6)
	Pub. goods	Debt	Gov. admin.	Pub. goods	Debt	Gov. admin.
Fractionalization	-394.531	-1493.369	13.850			
	(666.216)	(1945.131)	(44.190)			
Polarization				-196.938	-646.712	16.555
				(320.221)	(905.155)	(22.406)
City fixed effects	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Observations	1753	1753	1753	1753	1753	1753
R-squared	0.990	0.344	0.990	0.990	0.344	0.990

Robust standard errors (clustered at the council level) in parentheses. See text for a description of each expenditure category. *** p<0.01, ** p<0.05, * p<0.1

Figure 8 presents the first piece of evidence that an increase in the diversity of the governing body decreases expenditures on public goods. In Figure 8, each city's expenditure

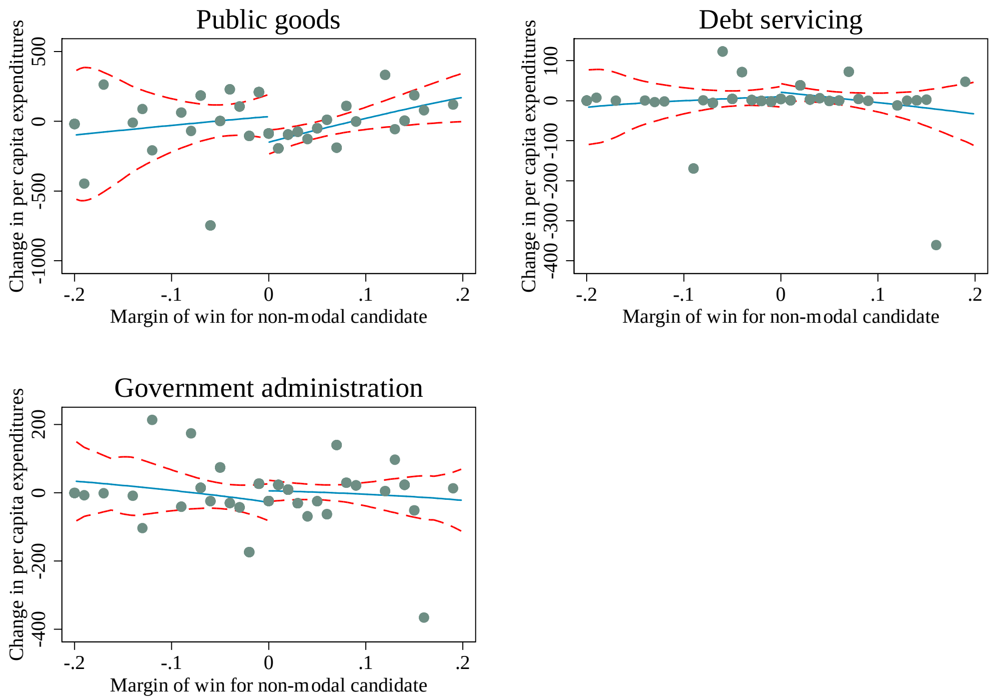
[53] Table 12 presents estimates from the sample of city-year pairs where the ethnicity for each councilmember is known. Restricting the sample to the set of cities that ever experience an election between a modal and non-modal candidate produces similar results.

categories are collapsed into two periods (before and after an election between a modal and non-modal candidate occurs). We then take the difference between those two periods, which allows us to analyze how average spending on public goods, debt, and government administration evolved following an election between a modal and non-modal candidate. On the x-axis, we organize cities by the non-modal candidate's margin of victory. Note that the margin of victory will be negative if the non-modal candidate lost the election. We then non-parametrically estimate the relationship between "non-modal win margin" and changes in per-capita expenditures. The first line was estimated for non-modal candidate losses (margins less than zero) while the second was estimated for non-modal wins (margins greater than zero). For ease of interpretation, we only display the mean change in expenditures across all elections in one-percent intervals – this is done to avoid displaying 150 observations, which produces a disorderly figure.

The top left panel of Figure 8 displays the change in per-capita expenditures on public goods. The break between the two lines that occurs at zero (i.e. when the non-modal candidate was narrowly elected) indicates that per-capita public goods expenditures fell in the years following a non-modal election.[54] It is important to note that this result is obtained without adopting any controls. That is, we are only analyzing a simple difference between pre and post-election expenditures. Notice, however, that as the election becomes less random (i.e. the non-modal candidate wins by a margin of more than ten percent) the change in public goods expenditures returns to zero. This highlights the importance of using quasi-experimental methodology to deal with endogeneity. We repeat this exercise for debt expenditures (top-right panel) and government administration (bottom panel), but find no evidence that either of those expenditure categories are affected by an increase in diversity.

[54] The dashed lines represent the 95-percent confidence interval.

Figure 8: Local polynomial smooth estimates of the change in per capita expenditures following the election of a non-modal candidate



Sample restricted to the set of cities that ever experience an election between a modal and non-modal candidate. Each point represents the average change in expenditures for each percentage point interval. The dashed lines represent the 95-percent confidence interval.

Table 13 reports the results of our regression discontinuity estimates. We interact the indicator "non-modal wins" with various functional forms for "margin of victory," but these coefficients are not displayed.[55] Ultimately, we find results consistent with the information displayed in Figure 7. As we saw in the previous section, the election of a non-modal candidate leads to higher diversity on the council. Here we find that this increase in diversity is then associated with a reduction in per-capita public goods expenditures on the order of \$150 to \$180, depending on whether "margin of victory" is modeled as a second, third, or fourth-degree polynomial. Cities appear to shift their spending away from public good expenditures, which would immediately benefit the population, and towards debt servicing and government administration, but the standard errors on these estimates are large. These results are most

[55] Full results available upon request.

consistent with the Alesina et al. (1999) argument that diversity in a group generates disagreement over public goods, which in turn implies a reduction in the willingness to spend.[56]

Table 13: The impact of a non-modal victory on government spending per capita (Regression discontinuity approach)

Outcome (cutoff margin)	(1) Pub. good	(2) Debt	(3) Gov. admin
Non-modal wins	-147.072***	16.37	3.903
(2^{nd} degree poly.)	(51.658)	(14.426)	(17.483)
Non-modal wins	-183.105***	19.777	10.083
(3^{rd} degree poly.)	(57.200)	(16.411)	(18.800)
Non-modal wins	-183.022***	16.033	4.439
(4^{th} degree poly.)	(64.849)	(17.954)	(22.308)

Robust standard errors (clustered at council-level) in parentheses. See text for a description of each expenditure category. *** p<0.01, ** p<0.05, * p<0.1

2.5.c An instrumental variables approach

Table 13 documents that a close non-modal victory has a negative impact on public good spending. Keep in mind, though, that we use "non-modal" wins as a reduced form proxy for an increase in continuous measures of diversity (fractionalization and polarization). Before proceeding to some additional probing of our main result, we aim to document that the results in Table 13 are driven by a change in continuous diversity measures. In this section we adopt an instrumental variables strategy that essentially uses the regression-discontinuity approach as an instrument for diversity (as measured by fractionalization or polarization), and then estimate the impact of *predicted* diversity (i.e. predicted fractionalization or predicted polarization) on government spending. Specifically, we estimate the following equations:

[56] It is possible that the decrease in spending we observe reflects the fact that budgets in treated cities are not growing as fast as they are in untreated cities. Consistent with this claim, Hopkins (2009) shows that diverse cities are less likely to raise taxes. Nevertheless, we find that an insignificant relationship between a non-modal win and per-capita revenues. These results are available upon request.

First stage:

$$Diversity_{ct} = \propto + \beta_1 \mathbf{1}[Non-modal\ wins_{ct}] + \beta_2 F(win\ margin_c) + \beta_3 \mathbf{1}[Non-modal\ wins_{ct}] * F(win\ margin_c) + year\ FE's + city\ FE's + \varepsilon \quad (2.4)$$

Second stage:

$$y_{ct} = \propto + \beta_1 \widehat{Diversity}_{ct} + year\ FE's + city\ FE's + \varepsilon \quad (2.5)$$

where the first stage mirrors the standard parametric regression-discontinuity approach used in Equation (2.3) and $Diversity_{ct}$ will either reflect the degree of fractionalization or polarization within city council c at year t. As in the previous section, y_{ct} will capture either: per-capita public good spending, per-capita debt servicing, or per-capita spending on government administration. Standard errors are clustered at the city council level.

Table 14: Two-state least squares estimates of the impact of diversity on government spending per capita

	(1) Pub. goods	(2) Debt	(3) Gov. admin.	(4) Pub. goods	(5) Debt	(6) Gov. admin.
Fractionalization	-2105.060**	248.695	96.793			
	(907.048)	(186.704)	(205.488)			
Polarization				-1218.116**	143.669	56.367
				(524.425)	(105.787)	(120.369)
City fixed effects	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y
Observations	503	503	503	503	503	503
R-squared	0.916	0.626	0.764	0.915	0.625	0.765

Robust standard errors (clustered at the council level) in parentheses. See text for a description of each expenditure category. *** p<0.01, ** p<0.05, * p<0.1

Table 14 reports our second-stage results for each of the three major expenditure categories.[57] Again, we find that greater diversity in the council generates a significant reduction in the per-capita spending on public goods. To interpret this coefficient, note that the election of a non-white to a five-person and all-white council would increase fractionalization from 0 to 0.4. If this were the case, our results suggest that per-capita spending on public goods would decline by approximately \$842. This is essentially an upper bound on the tangible impact of our results,

[57] Table 14 presents results using a third-degree polynomial to model “win margin”. Using a second-degree of fourth-degree polynomial presents nearly identical results. These results are available upon request.

as the election of a non-white to an otherwise all-white council is the largest possible marginal change in fractionalization.

2.6 ADDITIONAL RESULTS

In the previous section, we showed that the election of a candidate whose ethnicity differs from that of the modal ethnicity increases the diversity of the city council. Moreover, we found that increase in diversity to result in less spending on public goods. Having documented our main result (and the fact that the fall in public goods spending operates through a change in the continuous diversity measures), we now further consider the robustness of these results.

First, in the previous section we collapsed a number of areas of spending into one broad "public good" category. This was done largely because cities, especially in California, vary in the public goods they do or do not provide.[58] Thus, collapsing into a single "public good spending" measure allows for a uniform treatment of cities in the sample. There is some question, though, as to whether our documented result is truly a decline in public good spending generally, or whether it is driven by one specific category. In Table 15 we unbundle public goods expenditures to analyze the impact of diversity on the following expenditure categories: community development, culture and leisure, health, public safety, and public transit. As in Table 13, we only report the coefficient for "non-modal wins" and each row reports this coefficient from a regression that models "margin of victory" as either a second, third, or fourth degree polynomial. Although the standard errors are larger than in our main results, the results are consistently negative for all categories of public goods.

Next, we consider whether our results are driven by diversity per se, or whether they are driven by a change in minority representation. This requires further probing, as it is true that a "non-modal" candidate will often be an ethnic minority. The empirical finding would be of interest either way, but our hypothesized mechanism (disagreement driven by multiple opinions on the council, leading to a decline in spending) is most plausible if the result is driven by

[58] It is common for cities to make contracts with private entities or other governments (either a nearby city or the county) to provide some services.

diversity. To ensure that these results are not simply driven by the occasional election of a minority in an otherwise mostly-white council, we drop all situations where white is the modal race. This leaves us with 323 council-year pairs. Again, we find a large and statistically significant decrease in per-capita public good expenditures resulting from the election of a "non-modal" candidate. The results presented in Table 16 indicate that our earlier findings are not being driven by the election of a minority to a mostly white council. For these cities, however, it appears that the election of a non-modal candidate also reduces expenditures on government administration.

Table 15: Impact of diversity decomposed by different categories of public spending per capita

Outcome (cutoff margin)	(1) Community dev.	(2) Culture/ leisure	(3) Health	(4) Public safety	(5) Public transit
Non-modal wins	-29.710	-14.840	-45.153*	-13.318	-26.650
(2^{nd} degree poly.)	(24.212)	(15.176)	(24.143)	(9.499)	(18.322)
Non-modal wins	-36.075	-18.020	-56.026**	-20.497*	-22.451
(3^{rd} degree poly.)	(27.565)	(16.928)	(26.200)	(11.406)	(19.572)
Non-modal wins	-91.794***	4.193	-51.801*	-14.385	-1.595
(4^{th} degree poly.)	(34.119)	(20.710)	(27.607)	(12.631)	(21.800)

Robust standard errors (clustered at council-level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

In a related test, we examine the interaction between our result and the diversity of the city. By partitioning our sample, we address the question of whether a more diverse council reduces spending only when they represent a relatively homogenous city. In Table 17, we partition the sample by city-level diversity. Specifically, we estimate Equation (2.5) using a third order polynomial for $F(margin\,of\,victory)$ on two samples – those with city-level fractionalization less than the median level of city-fractionalization within the sample (0.57) and those with city-level fractionalization greater than the median. Whether the diversity of the city is above or below the median, the election of a non-modal candidate has (statistically) the same affect on public goods expenditures. Interestingly, it appears that spending on government administration falls in less diverse cities while spending on debt servicing falls for more diverse cities.

Table 16: The impact of a non-modal victory on per-capita spending – removing cities where "White" is the modal race

	(1)	(2)	(3)
Outcome (cutoff margin)	Pub. good	Debt	Gov. admin
Non-modal wins	-200.165***	6.827	-24.069
(2nd degree poly.)	(63.919)	(8.420)	(15.222)
Non-modal wins	-235.812***	3.253	-29.735**
(3rd degree poly.)	(72.825)	(7.829)	(13.937)
Non-modal wins	-271.148***	1.469	-45.468**
(4th degree poly.)	(88.459)	(8.730)	(18.391)

Robust standard errors (clustered at council-level) in parentheses. See text for a description of each expenditure category. *** p<0.01, ** p<0.05, * p<0.1

Table 17: The impact of a non-modal victory on per-capita spending – Partitioning the sample based on city-level diversity

	(1)	(2)	(3)
Outcome (sample restriction)	Pub. good	Debt	Gov. admin
Non-modal wins	-162.313*	19.501	-30.274*
(if city fract<0.57)	(84.952)	(23.628)	(16.768)
Non-modal wins	-182.392*	-33.485*	53.245
(if city fract>0.57)	(100.925)	(17.161)	(66.622)

Robust standard errors (clustered at council-level) in parentheses. See text for a description of each expenditure category. *** p<0.01, ** p<0.05, * p<0.1

Table 18: The impact of diversity on per-capita spending over time – RD approach (3rd degree polynomial)

	(1) Pub. goods	(2) Debt	(3) Gov. admin.
Non-modal wins	-201.039***	16.194	4.155
	(66.426)	(13.810)	(22.500)
Non-modal wins * After 1st year	-50.541	18.044	16.476
	(107.532)	(14.166)	(33.555)
City fixed effects	Y	Y	Y
Year fixed effects	Y	Y	Y
Observations	562	562	562
R-squared	0.940	0.649	0.766

Robust standard errors (clustered at council-level) in parentheses. See text for a description of each expenditure category. *** p<0.01, ** p<0.05, * p<0.1

Next we seek to understand how our main result evolves over time. Is the decline in spending a temporary shock that is quickly remedied as the council discovers ways to circumvent disagreement (i.e. bundling public goods, or favor trading)? We assess this by interacting our main treatment variable (and all interactions with margin of victory) with an indicator variable called "After 1st year".[59] This variable equals zero if the expenditure data correspond to the first year that the non-modal candidate could potentially impact the budget. The variable is set to one afterwards. It is important to note that our panel is relatively short, and so for most of the sample "After 1st year" simply implies "in the second year". Estimates interacting "Non-modal wins" with "After 1st year" are presented in Table 18. Keeping in mind that the impact of "Non-modal wins" and "Non-modal wins * After 1st year" are additive, we find no evidence of a reversal of the gridlock effect. If anything, public good spending falls even more as the non-modal candidate becomes more established within the council.

[59] Because we have a short unbalanced panel, we cut the data too thin when we attempt to examine dynamics in any greater detail.

2.7 PROBING THE MECHANISM

Thus far, we have seen that diversity in a city council leads to a reduction in spending on public goods. This result is consistent with the argument that more diversity within a council leads to disagreement and "gridlock." However, there are of course other potential explanations. While we cannot directly observe gridlock, in this section we offer two additional tests to further probe the mechanism driving our main result.

2.7.a Members of diverse councils are less likely to be reelected

One alternative explanation for why spending falls is that diverse councils find ways to provide the same public goods more efficiently, which would be a positive outcome for the city. To assess whether a decline in spending is perceived as a negative outcome by voters (e.g., inaction due to an inability to agree on public good provision, rather than government eliminating wasteful spending), we turn to a measure of voter satisfaction: the electoral success of city council members in the *next* election that they face. That is, we observe that a council in City A experienced a shock to diversity, while a council in City B did not. We also observe the election returns for all members of City A's council and City B's council when they run for re-election (if they do so). We examine whether the share of the vote received by City A incumbents suffers relative to City B incumbents as a result of the shock to diversity.[60]

We again employ a regression-discontinuity approach, taking the election of a non-modal candidate as a shock to diversity. Here, the unit of analysis is the council member, and the outcome variable is that council member's share of the vote. The data are setup as a panel where we observe each candidate's vote share twice. The first vote share comes from the election that brought them into office and the second vote share comes from the election following the potentially treatment-inducing election. This allows us to control for candidate ability by including candidate fixed effects. Because the number of candidates in the race mechanically impacts vote share, we include indicator variables that account for the number of candidates

[60] Because councilmembers can choose whether or not to run for re-election, it is perhaps important to note that incumbents ran for re-election 43.4 percent of the time following a non-modal win and 45.5 percent of the time when the non-modal candidate lost. Thus, selection issues are not a major concern.

seeking election. As before, we also include year fixed effects and cluster standard errors at the council level.

Table 19: The impact of a non-modal win on city council members' electoral success

	(1) Vote share (2nd degree poly.)	(2) Vote share (3rd degree poly.)	(3) Vote share (4th degree poly.)
Non-modal wins	-0.038**	-0.050***	-0.029
	(0.015)	(0.017)	(0.020)
Year fixed effects	Y	Y	Y
Candidate fixed effects	Y	Y	Y
FE's for num. of candidates in race	Y	Y	Y
Observations	798	798	798
R-squared	0.957	0.957	0.958

Robust standard errors (clustered at the council level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Results are reported in Table 19. Columns 1, 2, and 3 report the same specification using second, third, and fourth degree polynomials respectively. In all specifications, there is evidence that voters are less likely to re-elect a councilmember who experienced an exogenous increase in diversity. Specifically, we find that the vote share for a treated council-member drops by 3 to 5 percentage points (albeit with some lost precision in the fourth-degree polynomial specification). This result is not simply driven by the non-modal candidate as removing the winner of the potentially treatment-inducing election produces nearly identical results.[61] Based on these results, it seems that the decline in public good spending is indeed an outcome that is viewed as dissatisfactory to voters.

2.7.b Diversity impacts spending most in segregated cities

In order for gridlock to explain our results, it must be the case that there is disagreement between councilmembers of different ethnicities. While we have no direct evidence on whether this is or

[61] These results are available upon request.

is not the case, we explore one potential source of disagreement that is likely to be prominent (and does not rely on intrinsic differences in preferences across groups). In particular, for public goods or services that are somewhat local in nature, councilmembers may have a preference for positioning the public good so that it is accessible to their own group.

For example, suppose a city government is considering building a park (or some other public good that will inevitably benefit nearby citizens more than others). While there may be no difference across ethnic groups with regards to the *type* of park that is built, councilmembers may aim to position the park as close to members of their own ethnic group as possible. Note, however, that in order for this to be a source of disagreement, there must be some degree of segregation within the city.

Thus, if our results are driven by disagreement (and if the location of public goods is one prominent source of disagreement), then we should expect our results to be stronger in cities that are more segregated. To test this, we use tract-level census data to construct a measure of segregation for each city in the sample. In particular, we measure segregation using the *multi-group dissimilarity index,* as proposed by Reardon and Firebaugh (2002).[62] The index runs from 0 to 1, where 0 indicates that a city is totally integrated and segregation increases as the index approaches 1.

We split our sample at the median level of segregation according to the multi-group dissimilarity index (in our sample, the median is 0.22), and rerun our main specifications separately for highly segregated cities and other cities. Results are reported in Table 20. Columns 1-3 report results for highly segregated cities. It is indeed the case that results are stronger in highly segregated cities, both with respect to magnitude of coefficients and statistical precision. While we can not claim anything definitive about the mechanism, this result is also consistent with gridlock, as preference for making (local) public goods accessible to one's own group is likely to be an important source of disagreement.

We should also note that our result here is roughly consistent with (and may help speak to mechanisms driving) Alesina and Zhuravskaya's (2011) results. In a cross-country comparison, they find that ethnic segregation is associated lower quality of government along a number of

[62] This index is an extension of the dissimilarity index, which was proposed by Duncan and Duncan (1955) and is commonly used within economics to measure segregation (see, for instance, Cutler, Glaeser, Vigdor, 1999; Ananat, 2006; and Baum-Snow and Lutz, 2011). Recently, La Ferrara and Mele (2007) use the multi-group dissimilarity index to explore the relationship between racial segregation and public school expenditures.

dimensions including "government effectiveness", which captures citizen satisfaction with government provision of goods, services, and infrastructure.

Table 20: Impact of diversity on spending, decomposed by segregation of city

	Above median segregation			Below median segregation		
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome (cutoff margin)	Pub. good	Debt	Gov. admin	Pub. good	Debt	Gov. admin
Non-modal wins	-201.338***	6.323	40.938	-55.832	34.420	-26.412
(2nd degree poly.)	(62.782)	(14.615)	(34.120)	(96.070)	(33.944)	(18.692)
Non-modal wins	-241.953***	-2.710	17.787	-55.703	40.462	-22.618
(3rd degree poly.)	(71.307)	(12.533)	(41.574)	(106.424)	(39.780)	(18.202)
Non-modal wins	-199.359**	-17.230	11.743	-58.387	36.786	-13.027
(4th degree poly.)	(80.554)	(14.966)	(49.740)	(106.980)	(40.337)	(21.412)

Robust standard errors (clustered at the council level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

2.8 CONCLUSION

We analyze the relationship between ethnic diversity and public good provision by constructing a novel dataset linking the ethnicity of city councilmembers to election outcomes and expenditure decisions. This allows us to exploit close elections as a source of random variation in the ethnic composition of a city council. We first show that the narrow election of a candidate whose ethnicity is *not* the city's modal ethnicity is a plausibly exogenous shock to diversity within the city council. More precisely, the election of a non-modal candidate is associated with an increase in diversity at the government level is not associated with systematic differences in other city-level characteristics (specifically, city-level ethnic diversity or city-level political composition). We then implement a regression discontinuity design, which allows us explore how this increase in diversity affects the provision of public goods.

Results suggest that increases in diversity lead to gridlock; cities reduce the amount they spend on public goods. We find no evidence to suggest that this effect disappears over time, which indicates that cities are not circumventing disagreement through the use of logrolling and favor trading. The results are strongest in cities with high levels of segregation. In cities with high ethnic segregation, the physical location of (very local) public goods offers a source for disagreement (even where there is no other ethnic disagreement over "type" of public good being offered). Thus, we take this as suggestive of the gridlock hypothesis. Moreover, we find that members of diverse councils face diminished reelection prospects, ruling out the possibility that the drop in spending is perceived as a positive outcome by citizens.

These results reconcile a debate within the literature. Some have argued that increases in diversity will lead to disagreement over the type of public good provided and a reduction in spending, what we call the "gridlock" hypothesis. Empirical support for this claim is mixed, but previous studies focus on diversity within the city. This makes it difficult to identify a source of quasi-random variation, but more importantly, focusing on within-city diversity ignores the fact that diversity is filtered through the political process. Recent empirical work questioning the validity of the median voter theorem in a representative democracy indicates that the diversity of the governing body might be a more appropriate unit of analysis. Furthermore, if the median voter theorem does not hold, theoretical ambiguity as to how diversity affects decision-making in a small-group setting might explain the mixed results in the previous literature. For these reasons, we analyze the important yet unexplored question of how diversity within government affects public good provision.

Consistent with Boustan et al. (2013) and Hopkins (2011) we find that a simple fixed-effects analysis reveals no statistical relationship between ethnic diversity and public goods expenditures. However, the true relationship may be confounded. For instance, it could be the case that as a city becomes increasingly diverse and elects a diverse government, it also experiences population growth that requires an increase in infrastructure investment. In this case, the increase in public good expenditures resulting from population growth might confound the effects of diversity. Our empirical strategy, which is not subject to these same critiques, indicates that diversity does in fact reduce public good expenditures, which is consistent with the argument laid out in Alesina et al. (1999).

We should add several caveats to the interpretation of our main result. Most importantly, we emphasize that we are examining the impact of diversity on a single outcome, which certainly does not represent the entirety of government's influence on citizens' wellbeing. A bit more specifically, although *spending* falls, we are unable to say anything about actual provision of the public good. As noted, it is at least possible that spending falls because government is cutting wasteful or inefficient spending. Our results on subsequent electoral success of members of diverse councils suggest that this is not the case. It could be the case, however, that spending falls for services not valued by those from previously under-represented groups. To more fully assess the welfare and equity implications of this decline in spending, future research should find group-specific outcomes to assess whether certain groups disproportionately benefit or suffer from a more diverse government. Data on consumption or enjoyment of publicly provided services, for instance, would allow future researchers to more fully address these issues.

3.0 TYPHOID FEVER, WATER QUALITY, AND HUMAN CAPITAL FORMATION

Investment in water purification technologies led to large mortality declines by helping eradicate typhoid fever and other waterborne diseases. This paper, co-authored with Joe Ferrie, Martin Saavedra, and Werner Troesken, seeks to understand how these technologies affected human capital formation. We use typhoid fatality rates during early life as a proxy for water quality. To carry out the analysis, city-level data are merged with a unique dataset linking individuals between the 1900 and 1940 censuses. Parametric and semi-parametric estimates suggest that eradicating early-life exposure to typhoid fever would have increased earnings in later life by 1% and increased educational attainment by one month. Instrumenting for typhoid fever using the typhoid rates from cities that lie upstream produces similar results. A simple cost-benefit analysis indicates that the increase in earnings from eradicating typhoid fever was more than sufficient to offset the costs of eradication.

3.1 INTRODUCTION

During the late nineteenth and early twentieth century, the waterborne disease that posed the most serious threat to American populations was typhoid fever—as of 1900, probably one of every three Americans would have contracted typhoid at one point in his or her life.[63] Typhoid was caused by the bacterium *Salmonella typhi*, and was typically contracted by drinking water tainted by the fecal wastes of infected individuals. A common transmission might have gone something like this. The family of a typhoid victim dumped the patient's waste into a cesspool or

[63] Troesken (2004).

privy vault. If the vault was too shallow or had leaks, it seeped into underground water sources. In turn, if these water sources were not adequately filtered, people who drew their water from them contracted typhoid. Typhoid rates in a given city or region were, therefore, highly correlated with the quality and extensiveness of water and sewerage systems.[64]

Before water filtration, typhoid fever was a major source of morbidity and mortality in American cities. Although typhoid only killed about 5-10 percent of those infected, those who survived were left more susceptible to other diseases. Consistent with this, a large economic literature has shown that water purification has large and diffuse health effects, accounting for roughly fifty percent of the decrease in U.S. mortality between 1900 and 1950 (Cutler and Miller 2005; Ferrie and Troesken 2008). While the extant literature has done a thorough job identifying and measuring the short-term health effects of improving water quality, economists have yet to identify the long-term economic effects of water purification. There is, in particular, no evidence on how exposure to waterborne diseases in childhood impairs human capital attainment twenty to thirty years later, nor is there any evidence regarding labor market outcomes.

Accordingly, the goal of our paper is to analyze the relationship between early-life exposure to typhoid fever and adult outcomes, particularly in terms of educational attainment and income. Although our explanatory variable of interest is local typhoid fever fatality rates during the neonatal, prenatal, and infant period, we interpret typhoid fever as a proxy for water quality for several reasons. First, typhoid was primarily spread by drinking water contaminated with the fecal matter of an infected individual. Consequently, typhoid fever epidemics occurred where the water was impure. For example, George F. Whipple argued that “the relation between [water quality and typhoid] is so close that the typhoid death-rate has been often used as an index of the quality of the water. Generally speaking . . . a very low death rate indicates a pure water, and a very high rate, contaminated water”. Second, cities eradicated typhoid fever by filtering and chlorinating water (Cutler and Miller, 2005). Third, bacteria counts do not exist for most cities during the study period. For these reasons, researchers have been using typhoid fever rates as a measure of water quality for at least a century.

To explore the relationship between early-life typhoid exposure and adult outcomes, we link city-year level typhoid fatality rates to children in the 1900 Census, which are then linked to

[64] This paragraph is based on George C. Whipple, *Typhoid Fever: Its Causes, Transmission, and Prevention*, New York: John Wiley & Sons, 1908, especially pp. 21-69.

adult outcomes in the 1940 Census. Parametric and semi-parametric results indicate that the eradication of typhoid fever would have increased educational attainment by one month and increased earnings by about one percent. Of course, one might be concerned that water quality is correlated with unobserved variables that might also influence human capital formation (e.g. schooling or other public investments). Given this, we implement an instrumental variables strategy. Because typhoid is a waterborne disease, cities that dump their sewage into a river will increase future typhoid rates for cities downstream. Using typhoid rates from the nearest upstream city as an instrument, we find results that are larger. Specifically, these results indicate that if typhoid had been eradicated, schooling would have increased by nine months and earnings would have increased by about nine percent. However, only the estimate for schooling is statistically significant. We also find some evidence that high typhoid rates during early-life impaired geographic mobility.

This paper complements the existing literature on the benefits to water purification. Cutler and Miller (2005) show that the adoption of water purification technologies decreased total mortality by 13 percent, infant mortality by 46 percent, and child mortality by 50 percent. Furthermore, those of lower socioeconomic status might have been the primary beneficiaries to water purification efforts. Troesken (2004) shows that water filtration reduced typhoid rates among African Americans by 52 percent, but reduced white disease rates by only 16 percent. Currie et al. (2013) analyze birth records and water quality in New Jersey from 1997-2007. They find that exposure to contaminated water during pregnancy is associated with lower birth weights and higher incidence of premature birth for the children of less educated mothers.

A growing literature has shown that early-life exposure to disease and deprivation has adverse effects on adult health and economic outcomes, lowering educational attainment, earnings, and mortality (see Almond and Currie (2011) for a detailed overview). Because the diseases that accompany contaminated water are manifold and often severe in their consequences, it is reasonable to hypothesize that early-life exposure to contaminated water will have long-run effects. Consistent with this literature, we find that exposure to contaminated water decreased educational attainment and earnings.

Our findings are particularly relevant for policymakers in the developing world. Many developing countries have yet to undertake efforts to purify their water, possibly because water purification is costly. Consequently, 780 million people do not have access to improved water sources, leaving them vulnerable to typhoid fever, cholera, and other waterborne diseases.[65] Each year 21.5 million persons contract typhoid fever while 5 million contract cholera.[66] Furthermore, diarrheal diseases alone account for 1.8 million deaths each year or 4.7 percent of deaths worldwide.[67] Cutler and Miller (2005) estimate the social return to water purification to be 23 to 1. Our results indicate that the discounted increase in earnings alone was sufficient to offset the costs of water purification.

3.2 TYPHOID FEVER

3.2.a. Living and dying with typhoid

Once they entered the body, typhoid bacilli had a one to three week incubation period. During incubation, an infected individual experienced mild fatigue, loss of appetite, and minor muscle aches. After incubation, the victim experienced more severe symptoms: chills, coated tongue, nosebleeds, coughing, insomnia, nausea, and diarrhea. At its early stages, typhoid's symptoms often resembled those of respiratory diseases and pneumonia was often present. In nearly all cases, typhoid victims experienced severe fever. Body temperatures could reach as high as 105º Fahrenheit.

Three weeks after incubation, the disease was at its worst. The patient was delirious, emaciated, and often had blood-tinged stools. One in five typhoid victims experienced a gastrointestinal hemorrhage. Internal hemorrhaging resulted when typhoid perforated the intestinal wall and sometimes continued on to attack the kidneys and liver. The risk of pulmonary complications, such as pneumonia and tuberculosis, was high at this time. The high fever associated with typhoid was so severe that about one-half of all victims experienced neuropsychiatric disorders at the peak of the disease. These disorders included encephalopathy

---

[65] Estimates for access to improved water sources taken from UNICEF: http://www.unicef.org/wash/index_watersecurity.html
[66] Typhoid and Cholera estimates from CDC: http://www.cdc.gov/nczved/divisions/dfbmd/diseases/typhoid_fever/technical.html and http://www.cdc.gov/cholera/general/index.html respectively.
[67] 1.8 diarrheal deaths from WHO: http://www.who.int/water_sanitation_health/diseases/burden/en/

(brain-swelling), nervous tremors and other Parkinson-like symptoms, abnormal behavior, babbling speech, confusion, and visual hallucinations. If, however, the patient survived all of this, the fever began to fall and a long period of recovery set in. It could take as long as four months to fully recover. Surprisingly, given the severity of typhoid's symptoms, 90 to 95 percent of its victims survived.[68]

That typhoid killed only 5 to 10 percent of its victims might lead one to wonder just how significant this disease could have been for human health and longevity. But typhoid's low case fatality rate understates the disease's true impact, because when typhoid did not kill you quickly and directly, it killed you slowly and indirectly.

A simple way to illustrate this last point is by looking at the results of a study conducted by Louis I. Dublin in 1915. Dublin followed 1,574 typhoid survivors over a three-year period. Comparing the mortality rates of typhoid survivors to the mortality rates of similarly-aged persons who had never suffered from typhoid, he found that during the first year after recovery, typhoid survivors were, on average, three times more likely to have died than those who had never been exposed to typhoid, and that in the second year after recovery, typhoid survivors were two times more likely to have died than non-typhoid survivors. By the third year after recovery, however, typhoid survivors did not face an elevated risk of mortality. The two biggest killers of typhoid survivors were tuberculosis (39 percent of all deaths) and heart failure (23 percent). Other prominent killers included kidney failure (8 percent) and pneumonia (7 percent).

Few studies on the effects of typhoid fever on pregnant women and infants exist before the invention of antibiotics, but case studies date back to at least the late 1890s. Hicks and French (1905) searched the medical literature for individual cases in which pregnant women were diagnosed with typhoid fever and the birth outcomes of the children were recorded. They found 30 such cases, most of which the infection started in the third trimester of the pregnancy. In about half of the births, typhoid bacilli were in the fetal blood or fetal organs. The fetuses with typhoid bacilli were delivered later into the infection (4.1 weeks from the onset of fever) when compared to fetuses that showed no signs of the bacilli (2.4 weeks). The contraction of typhoid fever during pregnancy also increased the risk of both miscarriage and pre-term delivery (Stevenson et al. 1951). Modern studies have found that pregnant women with typhoid fever who

[68] Whipple (1908), Curschmann and Stengel (1902, pp. 37-42), Sedgwick (1902, pp. 166-68). See also, Troesken (2004, pp. 23-36).

were treated with ampicillin, amoxicillin, or chloramphenicol typically did not have worse birth outcomes than women without typhoid fever (Riggall, Salkind, and Spizllacy 1974; Seoud et al 1988; Sulaiman and Sarwair 2007).

Sinha et al (1999) use blood tests from a large sample of households in India and show that typhoid fever is a common source of morbidity among children less than five years old. More recently, Case and Paxson (2009) present econometric evidence that early-life exposure to diarrhea and typhoid fever impairs cognitive functioning later in life. This finding is particularly important for the results presented in this paper, which show that increased exposure to typhoid as a child is associated with lower incomes and reduced educational attainment in adulthood. Along the same lines, Almond et al. (2012) and Costa (2000) show early-life exposure to disease can raise the probability of contracting diabetes, heart disease, and other chronic health problems later in life.

3.2.b. Typhoid as an indicator of water quality

In this paper, our primary indicator of water quality is typhoid fever. Before the advent of formal water testing, typhoid fever was taken as an indicator of water quality among public health experts. As George F. Whipple argued, "A very low [typhoid] death rate indicates a pure water, and a very high rate, contaminated water".[69] Similarly, a report on water quality in New York City in 1912 stated that "the death rate from typhoid fever is commonly taken as one index of the quality of a water supply."[70] This same report noted, however, that typhoid was an imperfect indicator of water quality because typhoid epidemics could sometimes be caused by milk, shellfish or other sources, and because the absence of typhoid did not guarantee the water in question was free from other pathogens that might cause diarrhea, cholera, or other diseases.

While it is true that typhoid could be spread by means other than water, in the era before water treatment those sources of infection accounted for only a tiny fraction of all typhoid outbreaks (Troesken 2004; Whipple 1908, pp. 131-33). In addition, as explained below, typhoid was eradicated not through shellfish inspection or milk pasteurization but through improvements in water quality. It is well established in historical demography that water filtration and chlorination were by far the most important in the eradication of typhoid (Ferrie and Troesken

[69] Whipple (1908), p. 228.
[70] *Engineering News*, May, 1913, p. 1087

2005; Melosi 2000, pp. 138-47; Troesken 2004). As for the idea that typhoid did not fully reflect all possible pathogens in the water, typhoid fever rates were correlated with the death rate from cholera and diarrhea (Fuertes 1897).

To demonstrate that typhoid fatality rates are correlated with water quality, we combine typhoid fatality data with water filtration data. 155 of the 204 U.S. cities with a population greater than 30,000 in 1915 had a municipally owned waterworks, and 73 of those waterworks employed some sort of filtration process (either sedimentation, coagulation, slow sand filtration, mechanical filtration, or chemical sterilization). The types of filtration technologies that these waterworks employed as well as the date that each technology was first implemented are reported in the 1915 General Statistics of Cities. These data are then paired with annual typhoid fatality rates. We are able to obtain annual typhoid data for 61 of the 73 "filtering" cities. Our typhoid data spans from 1880-1920 and was obtained from Whipple (1908) and various issues of the U.S. Mortality Statistics.

The timing variation of filtration adoption allows us to employ a differences-in-differences framework in order to study the extent to which filtration affects typhoid fatality rates. This is the same empirical strategy employed by Cutler and Miller (2005). In contrast to Cutler and Miller, our sample includes 61 cities and focuses on a slightly earlier time period (1880-1920 instead of 1900-1940). The benefit of focusing on an earlier time period is that the adoption of water filtration technologies is less likely to be confounded by other public health interventions, e.g. pasteurization. Our estimating equation is as follows:

$$ln(Typhoid_{ij}) = \alpha + \beta Filter_{ij} + \ city\ FE's\ + year\ FE's\ + \epsilon_{ij} \qquad (3.1)$$

where $Typhoid_{ij}$ is the typhoid fatality rate per 100,000 persons in city i during year j. The variable $Filter_{ij}$ is an indicator equal to one if city i has implemented some sort of filtration technology on or before year j. Standard errors are clustered at the city level.

Before turning to our results, it is important to note that cities often adopted more than one technology. For instance, a city might purify its water through the use of both mechanical filters and chemical sterilization. This is only a problem for our analysis if the technologies are adopted in different years. Of the 31 cities that employ more than one technology, 18 adopt their technologies in more than one time period. The question then becomes, which filtration date is the appropriate date to use? In the table below, we report results using the adoption of either the first or last filtration technology as the intervention date. Alternatively, we could: drop all cities that employ more than one technology; drop all observations that occur between the first and last intervention; or employ a series of "treatment" indicators (i.e. an indicator for adopting the first technology, an indicator for adopting the second technology, etc.). The results presented below are robust to each of these specifications.[71]

Table 21 presents the results from estimating equation (3.1). The first four columns use the first adoption date as the treatment date while the last four columns use the final adoption date as the treatment date. The first and fifth columns, which correspond directly to equation (3.1), indicate that typhoid fatality rates fell by approximately 20 percent following the adoption of water filtration technologies. In columns two and six we include city specific time trends to alleviate concerns that filtration technologies were adopted because of rising typhoid rates. In columns three and four, as well as seven and eight, we omit observations that occur after 1908. This is done to alleviate concerns about other public health initiatives, namely pasteurization. The year 1908 was chosen because it corresponds to the first citywide ordinance on pasteurization, which was implemented by Chicago. When post 1908 observations are omitted, we find that the adoption of water filtration technologies lowered typhoid rates by 25-50 percent depending on specification.

Table 21: The effect of water filtration on the ln(typhoid death rate)

	Using adoption of first filtration technology as intervention date				Using adoption of last filtration technology as intervention date			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
City begins filtering water	-0.1869**	-0.1922**	-0.3814***	-0.2477*	-0.2222**	-0.1649*	-0.4903**	-0.3519**
	(0.0880)	(0.0843)	(0.1414)	(0.1339)	(0.0884)	(0.0921)	(0.1995)	(0.1713)
City fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
City specific time trends		Y		Y		Y		Y
Sample restricted to before 1908			Y	Y			Y	Y
Observations	1871	1871	1101	1101	1871	1871	1101	1101
R-squared	0.7010	0.7871	0.6031	0.6952	0.7011	0.7860	0.5999	0.6948

Robust standard errors (clustered at the city level) in parentheses. Typhoid death rate is the number of typhoid deaths per 100,000. * $p<.1$ ** $p<.05$ *** $p<.01$

In Table 22 we show that water filtration also reduced the likelihood of extreme outbreaks. Throughout our entire sample the median typhoid fatality rate is 30 deaths per

[71] These results are available upon request.

100,000. The distribution of typhoid deaths is positively skewed as the mean typhoid fatality rate is 37 deaths per 100,000 and the maximum rate is 292 deaths per 100,000. In Table 22 we estimate a variation of equation (3.1) where the outcome variable is an indicator equal to one if the typhoid death rate for city i in year j is greater than either 50 or 76 deaths per 100,000. These rates correspond to the 75th and 90th percentiles, respectively. For this estimation, we use a probit variation of the preferred specification – omitting post 1908 observations and including city specific time trends.[72] The average marginal effects are reported in Table 22. The results indicate that the adoption of water filtration technologies decreased the likelihood of observing an epidemic greater than either 50 or 76 deaths per 100,000 by 20 to 30 percent.

Table 22: The effect of water filtration on typhoid death rates

	Using adoption of first filtration technology as intervention date		Using adoption of last filtration technology as intervention date	
	1[Typhoid rate>50 per 100k]	**1**[Typhoid rate>76 per 100k]	**1**[Typhoid rate>50 per 100k]	**1**[Typhoid rate>76 per 100k]
Average marginal effect from filtering water	-0.2055*	-0.2112***	-0.2853**	-0.3140***
	(0.1066)	(0.0609)	(0.1223)	(0.0625)
City fixed effects	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y
City specific time trends	Y	Y	Y	Y
Sample restricted to before 1908	Y	Y	Y	Y

Robust standard errors (clustered at the city level) in parentheses. * p<.1 ** p<.05 *** p<.01

3.2.c. The eradication of typhoid fever

For much of the nineteenth century, people believed typhoid arose spontaneously or spread through miasmas — miasmas were poisonous atmospheres thought to rise from swamps, decaying matter, and filth. In 1840, William Budd challenged these ideas, showing that typhoid spread through water and food. Budd recommended investment in public health infrastructure to halt the spread of typhoid. However, scientists who continued to espouse the idea that typhoid arose spontaneously, or spread through miasmas, vigorously attacked Budd and his new theory. Because of their attacks, Budd's recommendations were not soon implemented, and typhoid rates in Europe remained as high as 500 deaths per 100,000 persons. It took more than three decades

[72] Results are qualitatively identical if we run a linear probability model instead of a probit.

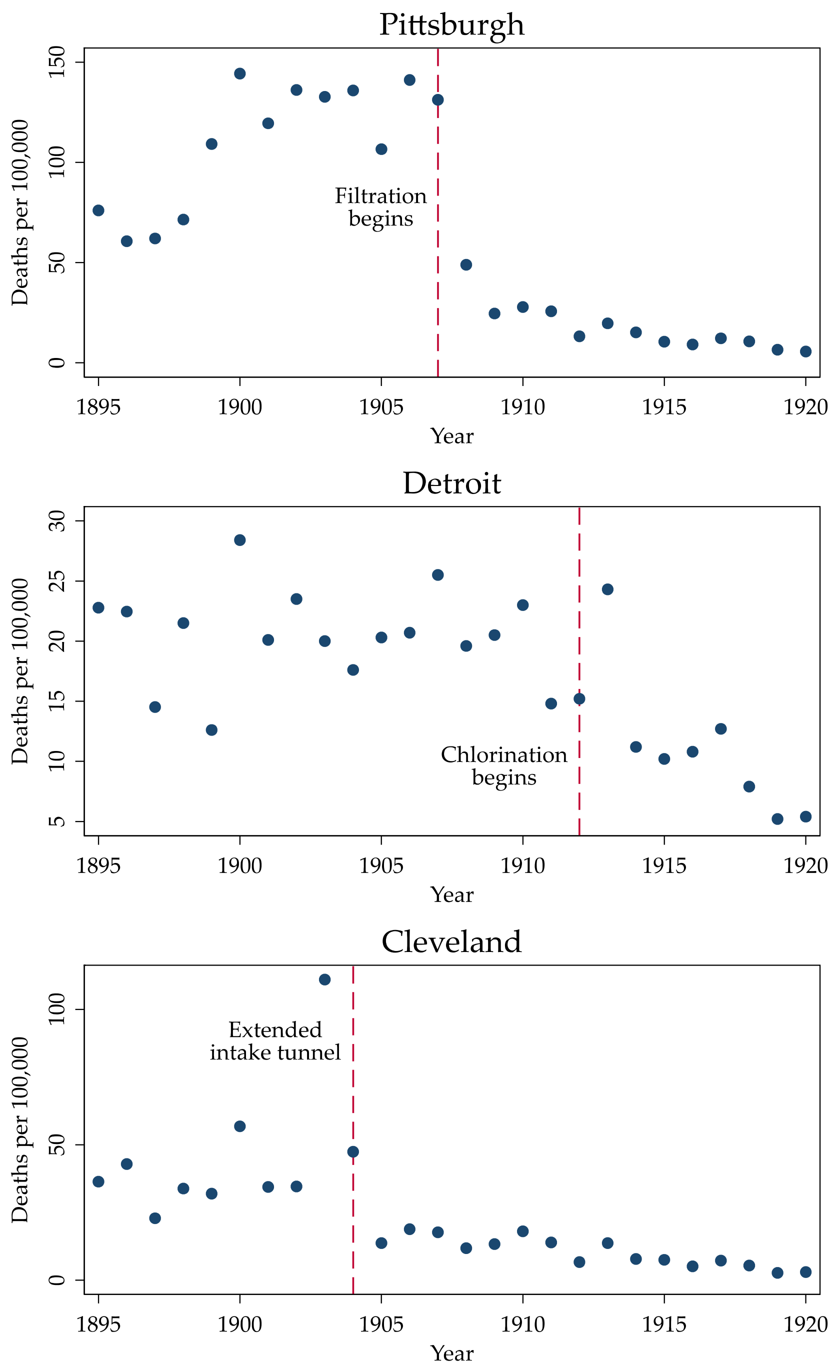
for Budd's theories and recommendations to take hold in England. In 1875, the British government passed the Public Health Act and began improving its public health systems. Ten years later, typhoid rates in England had fallen 50 percent.[73]

With the development of Budd's ideas in particular, and the germ theory of disease in general, public health officials in America and Europe came to agree: to control typhoid, cities needed to assure purity of drinking water through filtration and chlorination, and through sanitary sewage disposal. The experience of Pittsburgh, Pennsylvania highlights the effectiveness of water filtration in controlling typhoid fever. Pittsburgh drew its water from the Allegheny and Monongahela Rivers. Upstream from the city, seventy-five municipalities dumped their raw and untreated sewerage into the rivers, leaving Pittsburgh's typhoid rate higher than any other major U.S. city. Pittsburgh held this distinction throughout the late nineteenth century. Then, in 1899, Pittsburgh voters approved a bond issue for the construction of a water filtration plant. Unfortunately, political bickering delayed completion of the plant until 1907. Once the plant was in operation, though, typhoid rates improved, and by 1912, they equaled the average rate in America's five largest cities.[74]

As Figure 9 shows, in the years before the introduction of filtration, typhoid rates in Pittsburgh averaged about 100 deaths per 100,000. Within two years, filtration had reduced typhoid rates in Pittsburgh by roughly 75 percent. And through subsequent improvements and extensions in the city's water supply, typhoid rates were brought down to around 6 deaths per 100,000 by 1920. This represented a reduction of about 95 percent from pre-filtration levels. As impressive as the Pittsburgh example is, it represents a typical response of typhoid fever to filtration.[75]

[73] Budd (1918) and Melosi (2008, pp. 1-42; 60-61; and 110-13).
[74] For a survey of the effectiveness of water filtration (and other modes of improving water quality) in reducing typhoid rates, see Whipple (1908, pp. 228-66). On the Pittsburgh experience, see Troesken (2004, pp. 27 and 56).
[75] See Cutler and Miller (2005); Melosi (2008, pp. 136-48); Whipple (1908, pp. 228-72); Fuertes (1897); Sedgwick and MacNutt (1910).

Figure 9: Typhoid death rates



Data from Whipple (1908) and the 10th annual census report on mortality statistics.

Water filtration was not the only effective mechanism at decreasing typhoid fever. The other panels of Figure 9 show that typhoid fell following the introduction of chlorination in Detroit, and the extension of water intake cribs away from the shoreline in Cleveland. In all cities, the introduction of water purification technologies was followed by sharp reductions in the death rate from typhoid fever (Melosi, 2002; Ellms 1913; Cutler and Miller, 2005). The introduction of sewers also had an effect on mortality rates (Kestenbaum and Rosenthal, 2014; Beemer, Anderton, and Leonard, 2005; and Ferrie and Troesken, 2005). Although cities used many different technologies to purify their water, the majority of these interventions occurred after our sample period (1890-1900), which is why we use typhoid as a proxy for clean water rather than these technologies themselves. Specifically, only four cities for which we have typhoid data began filtering their water during the 1890s.

Some observers have argued that just looking at typhoid, as we have done here, understates the benefits of water filtration because eradicating typhoid has broad benefits. Specifically, eradicating typhoid affected mortality from a broad class of diseases and illnesses. The non-typhoid death rates that were the most responsive to improvements in water quality were infantile gastroenteritis (diarrhea), tuberculosis, pneumonia, influenza, bronchitis, heart disease, and kidney disease.[76]

The experience of Chicago nicely illustrates how improving water quality not only reduced deaths from typhoid fever but also a broad class of diseases not usually considered waterborne. From the late-nineteenth century onward, Chicago's primary water source was Lake Michigan. Unfortunately, Lake Michigan was also frequently polluted with sewage, which carried disease-causing pathogens. This pollution occurred because for much of the nineteenth century the city dumped its sewage directly into the lake, or into the Chicago River which flowed into the lake. Over the course of the nineteenth and early twentieth century, Chicago took two important steps in trying to prevent fecal pollution from entering the city's water mains. The first step occurred in 1893, when the city opened the Four-Mile water intake crib, the Sixty-eighth Street water intake crib, and permanently closured all shoreline sewage outlets.[77] The second

[76] Cutler and Miller (2005), Sedgwick and MacNutt (1910), Ferrie and Troesken (2008).
[77] For these projects and dates, see Chicago Bureau of Public Efficiency (1917); and *The Daily Inter-Ocean* (Chicago), January 1, 1894, p. 13.

step occurred in 1917, when the city opened the Wilson Avenue water intake crib and completed its citywide chlorination of the public water supply.[78]

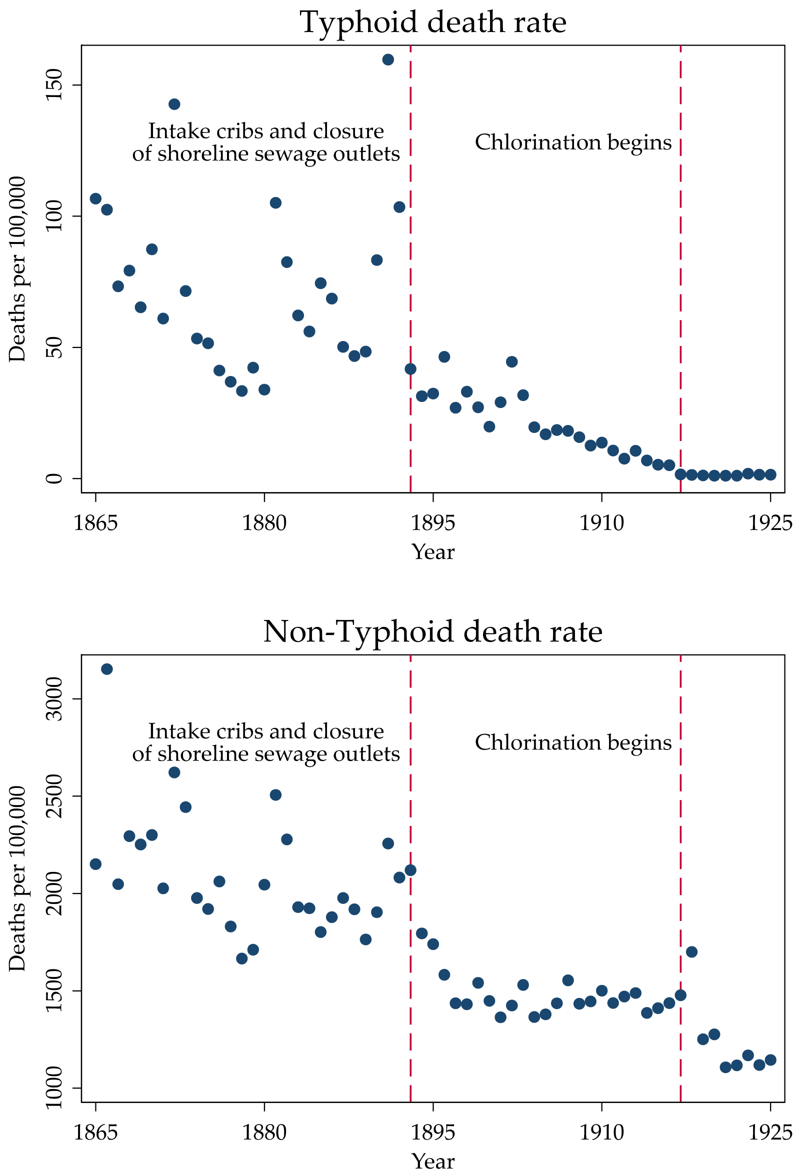
The completion of these projects corresponded with sharp drops in the city's death rate from typhoid fever. This can be seen in Figure 10, which plots typhoid rates in Chicago from 1865 to 1925. There are two vertical lines, each corresponding to the aforementioned technological improvements promoting water purity. Note in particular the dramatic effects of the Four Mile and Sixty-eight street water intake cribs and the closure of shoreline sewage outlets in 1893. Before 1893, typhoid rates averaged 73 deaths per 100,000, and death rates were often as high as 100 to 150. After 1893, death rates never rose above 50, and shortly after the opening of the Chicago drainage canal in 1900, rates never rose above 25. The installation and extension of chlorination around 1917 drove down typhoid rates still further until rates were hovering around 0 by the early 1920s.[79]

These improvements in water quality were also associated with large reductions in deaths from diseases other than just typhoid fever. This can be seen in the second panel of Figure 10, which plots the total death rate excluding deaths from typhoid fever. Again, the vertical lines correspond to the two regime changes in the city's water supply. The patterns are striking. Although death rates appear to be trending downward almost from the start of the time series, that trend is modest and highly variable. The two most prominent changes in the death rate are associated with improvements in the city's water supply. After the closure of shoreline sewage outlets and the opening of two new intake cribs in 1893, the total death rate quickly fell to 1500 per 100,000, and never again even remotely approached levels between 2000 and 2500, which were commonplace before 1893. Another sharp discontinuity is observed in 1917 when death rates fell to around 1100. The year 1917, moreover, coincides with the completion of the city's water chlorination system.

[78] See Cain (1977, pp. 57); *Municipal and County Engineering*, Vol. LVI, No. 1 (Jan.-June 1918), p. 6; Chicago Bureau of Public Efficiency (1917).

[79] The link between improvements in the city's water supply and reductions in typhoid rates did escape notice in the medical press. See the *Medical News*, November 21, 1896, p. 586; *The Daily Inter-Ocean* (Chicago), January 1, 1894, p. 13; and *Bulletin of the Chicago School of Sanitary Instruction* (Chicago Department of Public Health), Vol. XV, No. 9, Feb. 27, 1921, p. 34.

Figure 10: Death rates in Chicago



Data from Whipple (1908) and the 10th annual census report on mortality statistics.

One might argue that the decline in non-typhoid deaths was the result of other public health investments. However, Cutler and Miller (2005) show that death rates from pneumonia, diphtheria, and meningitis fell following the adoption of water purification technologies. Specifically, they estimate that for every one typhoid fever death prevented by water purification there were four deaths from other causes that were also prevented. Ferrie and Troesken (2008) present similar, though somewhat stronger, evidence along these lines. The available evidence suggests that these diseases improved with water filtration because typhoid was a virulent disease that left a person vulnerable to secondary infections even if he or she survived its direct effects.

3.2.d. Milk and typhoid fever

One concern with using typhoid to measure water quality is that typhoid was sometimes spread through mechanisms that do not, at least at first glance, appear to have been water-related. As already noted, typhoid could be spread through shellfish, individual human carriers, and milk. It is thus possible that declining typhoid rates not only reflect improvements in water quality but other environmental improvements as well, such as the introduction of pasteurization milk supplies in large cities, which commentators suggest had a large effect on diarrheal diseases, especially among young children (Rosenau 1912, pp. 189-229). If so, the results below would conflate the effects of improving water quality with better milk and overstate the significance of early-life water quality for later life economic outcomes.

In this section, we focus on the frequency and significance of milk-related transitions because milk was, after water, the most important source of typhoid outbreaks; transmissions through shellfish and typhoid carriers were comparatively rare events (see Whipple 1908, especially p. 271). As a preview, three conclusions emerge from the discussion that follows. First, although milk was the most important transmission mechanism after water, it accounted for a relatively small proportion of all typhoid cases in the years before widespread and effective water treatment techniques. Second, and along the same lines, pasteurization was probably the single most cited public health intervention for preventing milk-related typhoid transmissions (Whipple 1908, pp. 271-72). Yet, nearly all city-level mandatory pasteurization laws were passed after 1910, more than a decade after our sample period (1890-1900) ends. To the extent that pasteurization eliminated milk-related transmissions of typhoid, this treatment came too late to have driven the downward trends observed in our data. Third, in the era before pasteurization, most epidemics attributed to tainted milk actually originated polluted water sources. Put another way, milk-related typhoid epidemics were not typically milk-borne but ultimately started with impure water supplies. In this way, improving water supplies not only helped to eliminate the typhoid epidemics expressly designated as waterborne but also helped to combat milk-related outbreaks of the disease by making safer and cleaner the water used to dilute milk supplies and to clean milk containers.

To get a sense of how important milk was as a transmission mechanism during the late 1800s and early 1900s, we turn to several detailed city-level studies of typhoid infections. Between 1906 and 1909, the United States Marine Hospital Service (the forerunner to the U.S. Public Health Service) conducted an exhaustive survey to identify the origins of typhoid fever in Washington, D.C. The Marine Hospital Service concluded that for the years between 1906 and 1908 (inclusive), just under 11 percent of all typhoid cases in the city could be traced back to tainted milk supplies. A similar study conducted by the chief health officer of Richmond, Virginia concluded that out of roughly 2,300 cases of typhoid fever observed between 1907 and 1915, there was not a single case that could be attributed to milk. A broader study for the entire state of Virginia found that, at most, milk could be indicted in .8 percent of the all cases observed during the early 1900s (Frost 1916). More generally, nearly all observers believed that milk did not emerge as an important source of typhoid transmission in any given city until after that city had begun to clean up their water supplies; prior to water filtration, nearly all cases of typhoid were spread by bad water (Whipple 1908; Frost 1916; Ferrie and Troesken 2008; and Troesken 2004, pp. 22-33).

Whatever the frequency of milk-related typhoid epidemics, one of the most direct and effective ways of eliminating such epidemics was through pasteurization; boiling milk killed nearly all of the dangerous microbes, including typhoid, and forestalled the spread of disease. Most observers believed that pasteurization was the single most important factor in eradicating milk-related typhoid epidemics (Whipple 1908, pp. 271-78; United States Marine Hospital Service 1909, pp. 151-63). Chicago was the first city to pass an ordinance mandating pasteurization, and it did so in 1908. While a few other large cities soon followed suit, many cities did not have mandatory pasteurization as late as 1920: as of 1920, 9 of the 21 largest cities in the United States had no laws governing pasteurization, and pasteurization was even less common in smaller cities (American Public Health Association, 1920). In addition, there is some question as to whether city ordinances were mandating adequate treatment temperatures. For example, in New York City, health authorities required that pasteurized milk reach a temperature of 142 degrees for 30 minutes, despite the fact that the U.S. Public Health Service recommended that the milk reach a temperature of 145 degrees (see Public Health Reports, Vol. 41, No. 36, Sept. 3, 1926, pp. 1900-02).

Perhaps a central concern for our analysis is the relative roles played by pasteurization and chlorination. While previous research by Cutler and Miller (2005) and Ferrie and Troesken (2008) suggests that chlorination and improvements in water quality were far more important than pasteurization in driving down typhoid rates during this third regime, for the purposes of our analysis the question is a moot because both changes come after our study period. In particular, our analysis focuses on individuals born between 1890 and 1900, which precedes the advent of mandated pasteurization by roughly two decades. Our results, therefore, are mainly picking up the effects of filtration, not pasteurization and chlorination, which come much later.

The third point we wish to make is that a large proportion of typhoid epidemics identified as milk-borne were, in fact, waterborne. While the typhoid bacillus can survive in milk, cows do not carry typhoid and typhoid can only survive in human hosts. The question, therefore, is how does typhoid get into the milk in the first place if it does not come directly from the cow. There are two possible mechanisms. First, the cows and their milk might be handled by someone infected typhoid or by a typhoid carrier. Second, the cans and utensils used to distribute milk might have been washed with infected water, or in some cases, the milk might have been diluted with typhoid-polluted water. A 1909 report on milk and its relation to public health summarizes the causes of 138 milk-borne typhoid epidemics occurring between 1881 and 1907. One-hundred-and-nine of the 138 epidemics can be traced to a definitive single source; and 67 of these epidemics resulted from either the washing of utensils or the dilution of milk with infected well water. An additional six cases resulted from cows wading in sewage-polluted water. The remaining cases went unattributed as to origins or were attributed to workers that either experienced a bout of typhoid fever themselves or cared for someone infected with typhoid fever. Hence, of the 109 so-called milk-borne epidemics for which a definitive source could be identified, 67 percent originated from impure water sources or improper sewage disposal (United States, 1912).

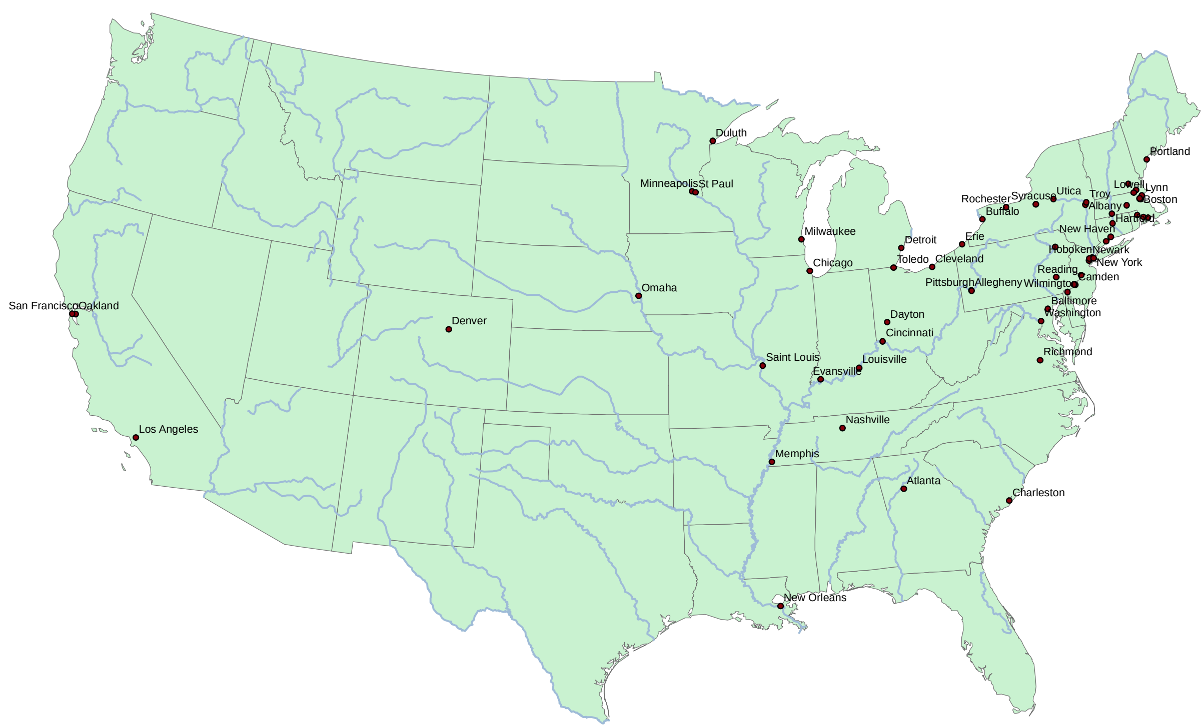
An example from Chicago highlights the importance of impure water in typhoid epidemics otherwise identified as milk-borne. In November of 1906, 22 cases of typhoid fever were reported in Chicago's Mont Clair neighborhood. The city quickly tested each of the surface wells and stopped the use of any wells that showed contamination. As summarized in the 1907 Chicago Health Report: "The presumption at the time was strong that the infection was carried in the water of several [infected] wells, as the samples of milk, ice, and city water examined were good." Further investigation revealed that each of the 22 cases received their milk from the same dealer and that, despite having access to city water, the milk dealer occasionally used well water to wash milk pails and utensils. That surface well was placed just 42 feet from an outhouse, which is viewed as the source of contamination. The Mont Clair outbreak nicely illustrates two points. First, milk and water were viewed as the most likely transmission mechanisms for typhoid fever. Second, even though the outbreak occurred through milk distribution, the ultimate cause of the epidemic was the use of typhoid-infected water. Having said all this, we do not wish to suggest that improvements in milk quality did not affect human health. It is widely appreciated that pasteurization played an important role in reducing infant mortality, primarily by eliminating exposure to bovine tuberculosis. The extent to which pasteurization matters for eradicating typhoid fever, however, is directly related to the purification (or rather, the lack of purification) of the water supply. Since pasteurization typically occurred after a city had taken steps to purify its water supply, the scope for pasteurization to affect typhoid fever was circumscribed.

3.3 DATA

Given the large literature showing how early-life exposure to disease impairs human capital formation, and given the observation that typhoid had large and diffuse health effects, one expects that typhoid would have also had large and diffuse effects on economic and social outcomes. To identify these effects we combine city-year level typhoid fatality data with a linked sample of males from the 1900 and 1940 censuses.

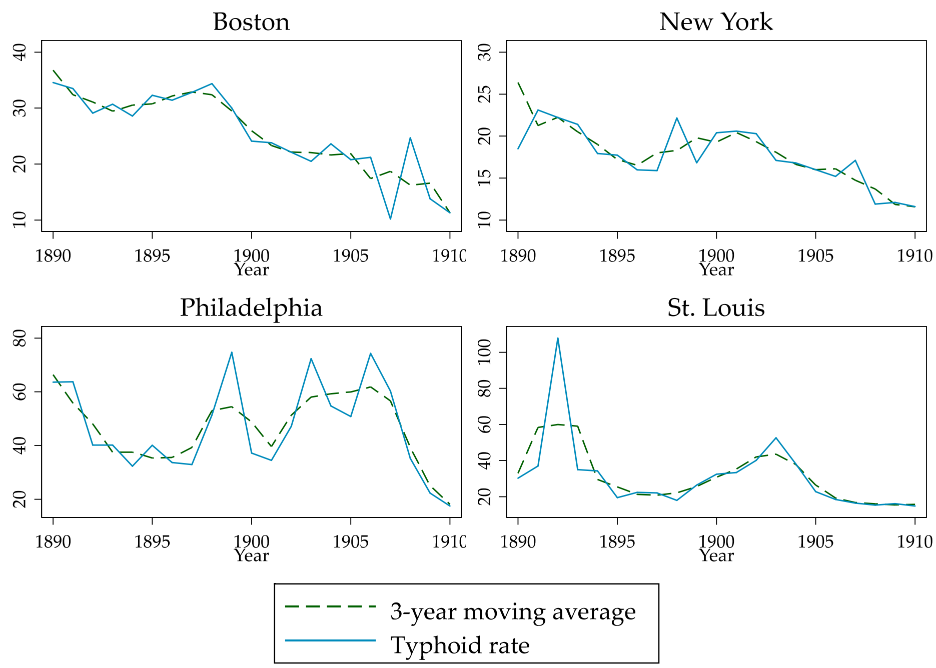
We obtain typhoid fatality rates in the late nineteenth and early twentieth centuries for 75 cities. This data was transcribed from Whipple (1908) as well as the 10th annual Census mortality statistics. Figure 11 maps the cities used in our analysis. These cities tend to fall within the top 100 in terms of population. In 1900, they had an average population of 225,364 and a median population of 94,969. The cities are predominantly located in the Northeast and the Midwest but include all regions of the continental United States.

Figure 11: Cities and rivers



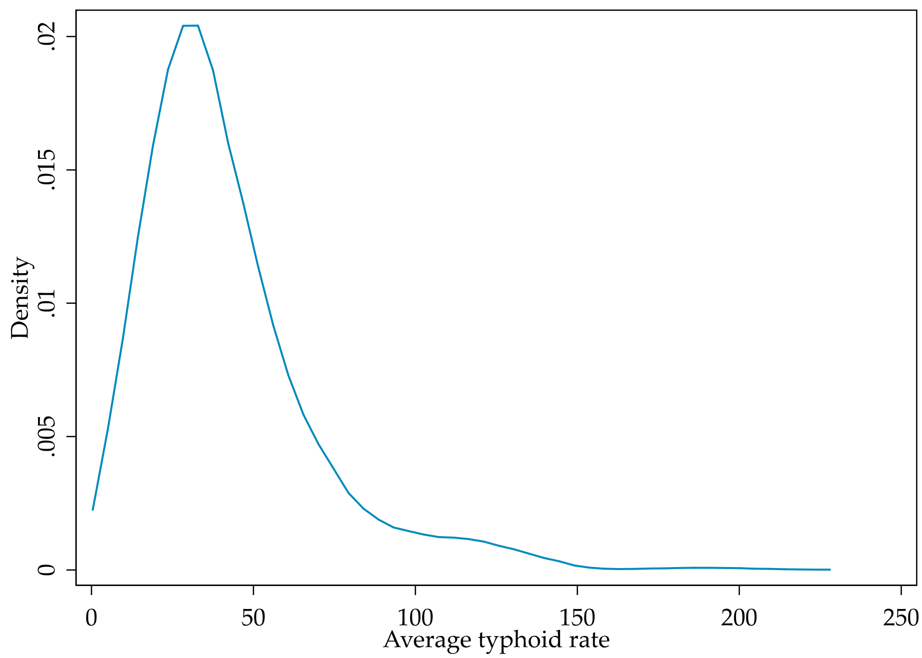
As a measure of early-life exposure to contaminated water, we average typhoid rates during the year of birth, the year before birth, and the year after birth. Figure 12 visually displays typhoid rates and the three-year moving average for Boston, New York, Philadelphia, and St. Louis between 1890 and 1910. Our analysis will focus on the three-year moving average. This has two advantages. First, because typhoid rates are volatile, the moving average provides a better proxy for average water quality. Second, the three-year moving average roughly corresponds with the prenatal, neonatal, and postnatal periods, which captures exposure during early life. Figure 13 plots the distribution of average typhoid rates during early life. The distribution is skewed right with a mean of 41.72 deaths per 100,000. The domain ranges from 10.39 deaths per 100,000 to 217.96 deaths per 100,000.

Figure 12: Typhoid rates



Data from Whipple (1908) and the 10th annual census report on mortality statistics. Typhoid fatality rate is the number of deaths per 100,000.

Figure 13: Distribution of typhoid rates during early life



Average typhoid rate during early life is the average typhoid rate during the year before birth, the year of birth, and the year after birth. The average typhoid fatality rate is the number of deaths per 100,000.

We merge this typhoid fatality data to linked micro data. This dataset links individuals observed in the 1940 and 1900 censuses that were born between 1889 and 1900. We restrict our analysis to males who, at the time of the 1900 census, were living in a city for which we have typhoid data. Because we treat the city of residence in 1900 as the birth city, we drop any individual that was born in a state other than their state of residence in 1900. We believe this assumption is reasonable given that the sample would be at most eleven years old in 1900.

Summary statistics are reported in Table 23. Age, education, income, homeownership status, and whether the individual moved from their birth city are taken from the 1940 census. These outcome variables are measured during peak earning years (ages 40-51). The percent of blacks is small because we are looking at individuals born in cities before the Great Migration. The average individual in our sample spent their early life in a city with an average typhoid rate of 42 deaths per 100,000.

Table 23: Summary statistics

	Mean	SD	Min	Max	Observations
Age in 1940	45.09	3.38	40	51	189,515
Education	9.47	3.16	0	17	184,331
Income	1511.99	1302.85	0	5001	176,821
Homeowner	0.48	0.50	0	1	153,932
Moved from birth city	0.61	0.49	0	1	189,515
Black	0.03	0.17	0	1	189,515
Birth order	2.97	2.12	1	91	189,515
Typhoid rate during early life	41.72	25.63	10.39	217.96	189,515

Age, education, income, homeowner status, and whether the individual moved from their birth city are taken from the 1940 census. Birth order was reported in the 1900 census. Although the maximum of 91 is likely a typographical error, the 99th percentile of birth order, which is 10, is plausible. Typhoid rate during early life is the average typhoid rate in the birth city from one year before birth, the year of birth, and one year after birth.

3.4 RESULTS

3.4.a. OLS results

In Table 24 we estimate the relationship between early-life typhoid and adult outcomes using the following equation:

$$y_{ijk} = \alpha + \beta Typhoid_{jk} + \gamma \mathbf{1}[black_i] + birth\ city\ FE's \qquad (3.2)$$
$$+\ birth\ year\ FE's + birth\ order\ \square E's\ +\ \epsilon_i$$

where the outcome for individual i born in city j during year k is either years of schooling, ln(income), homeownership status, or mover/stayer status in 1940. Typhoid is the average typhoid rate during early life for individuals born in city j during birth year k, where early life is defined as the year before birth until the year after birth. We cluster standard errors at the birth-city level. Each regression includes fixed effects for each birth city, birth year, and birth order. Because outcomes are taken from the 1940 census, controlling for birth year automatically controls for age. We find that typhoid during early life decreases educational attainment and adult income, but we find no effect on homeownership status or geographic mobility (mover/stayer status). These results indicate that if typhoid were eradicated, years of schooling would have increased by nearly one month and income would have increased by one percent.

Before probing the robustness of these results, it is important to note that a typhoid epidemic during early life could influence the distribution of adult outcomes in two ways. Typhoid fever negatively influences maternal and infant health. Accordingly, either the disease itself or the body's response to fight the infection may have had long-run effects on either adult health or cognition. This effect would have led to a "scarred" population and resulted in a negative correlation between early-life typhoid fever exposure and adult labor market outcomes. The alternative is that increased infant mortality during the epidemic led to a select sample of relatively healthier individuals. If survivorship bias is severe, then the surviving population during an epidemic, even if negatively affected by the epidemic, could be healthier than the population during a non-epidemic year. In this case, we would expect a positive correlation between early-life typhoid rates and adult labor market outcomes (Bozzoli, Deaton, and Quintana-Domeque 2007). In our data, we find that typhoid epidemics during early life are negatively correlated with labor market outcomes suggesting that the scarring effect dominates selection effect. Because typhoid increased infant mortality (Cutler and Miller 2005), our sample likely suffers from survivorship bias. This would imply that our estimates are underestimates of the scarring effect.

Table 24: The relationship between typhoid and adult outcomes

	Years of schooling	ln(income)	Homeowner	Mover
Average typhoid rate during early life	-0.0022***	-0.0003**	-0.0000	-0.0001
	(0.0005)	(0.0001)	(0.0001)	(0.0001)
Black	-1.7273***	-0.7135***	-0.2424***	0.0294
	(0.1722)	(0.0295)	(0.0197)	(0.0256)
Birth year fixed effects	Y	Y	Y	Y
Birth city fixed effects	Y	Y	Y	Y
Birth order fixed effects	Y	Y	Y	Y
The average effect from eradicating typhoid†	0.0912***	0.0128**	0.0000	0.0041
	(0.0227)	(0.0050)	(0.0027)	(0.0047)
Observations	184,331	141,857	153,932	189,515
R-squared	0.053	0.038	0.039	0.238

†The average effect from eradicating typhoid is calculated by multiplying the negative of the coefficient by the average typhoid rate during early life (41.72 deaths per 100,000). Robust standard errors (clustered at the city level) reported in parentheses. * p<.10; ** p<0.05; *** p<0.01

To better understand the extent to which selective mortality attenuates our results, we remove all cohorts whose average early-life typhoid exposure was greater than 99 deaths per 100,000 persons. Truncating the sample at the 95th percentile allows us to remove situations in which the selective mortality effect might dominate the scarring effect. OLS results with this sample restriction are presented in Table 25. Consistent with the attenuation argument described above, the effect of typhoid exposure is slightly larger when extreme epidemics are omitted. Specifically, we find that the eradication of typhoid fever would have increased educational attainment by one and a half months and increased earnings by about 2.5 percent. What is particularly surprising is how robust this finding is to our choice of cutoff. Truncating the sample at the 75th, 90th, or 95th percentile produces qualitatively identical results.[80] This suggests that selective mortality is only a problem during extreme epidemics, which is also consistent with typhoid fever's low case fatality rate.

Table 25: OLS results omitting cohorts whose early-life typhoid rate was greater than 99 deaths per 100,000

	Years of schooling	ln(income)	Homeowner	Mover
Average typhoid rate	-0.0034***	-0.0007***	0.0001	-0.0002
during early life	(0.0008)	(0.0002)	(0.0001)	(0.0002)
Black	-1.7054***	-0.7166***	-0.2398***	0.0274
	(0.1637)	(0.0285)	(0.0203)	(0.0263)
Birth year fixed effects	Y	Y	Y	Y
Birth city fixed effects	Y	Y	Y	Y
Birth order fixed effects	Y	Y	Y	Y
The average effect from	0.1271***	0.0247***	-0.0019	0.0060
eradicating typhoid†	(0.0316)	(0.0062)	(0.0049)	(0.0060)
Observations	174,919	134,764	145,975	179,921
R-squared	0.0536	0.0392	0.0405	0.2218

†The average effect from eradicating typhoid is calculated by multiplying the negative of the coefficient by the average typhoid rate during early life (37.46 deaths per 100,000). Robust standard errors (clustered at the city level) reported in parentheses. * p<.10; ** p<0.05; *** p<0.01

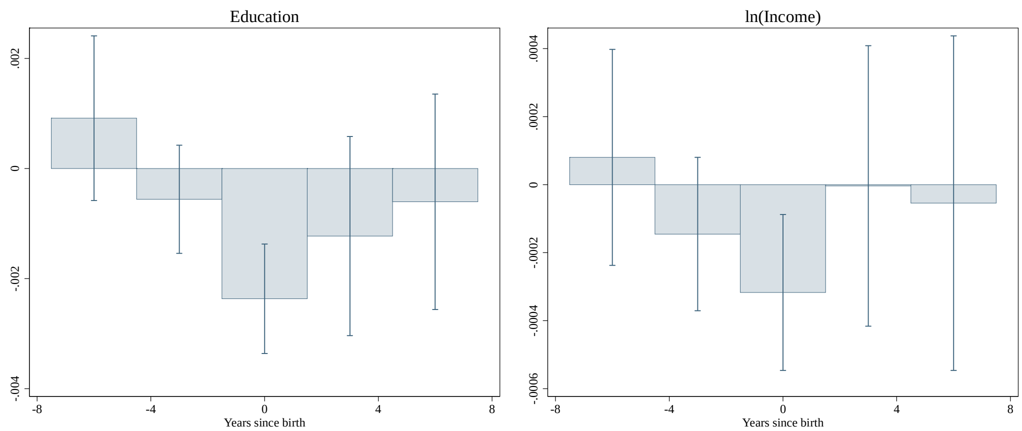
3.4.b. OLS robustness checks

To illustrate that these results are driven by early-life exposure and not exposure during other ages we estimate a variant of equation (3.2) that includes typhoid rates during the following years: 7 to 5 years before birth; 4 to 2 years before birth; 2 to 4 years after birth; 5 to 7 years after birth, as well as our measure of early-life exposure (1 year before birth to 1 year after birth). Figure 14 plots the 95 percent confidence interval for these estimates. Consistent with Table 24, Figure 14 illustrates that early-life typhoid rates are associated with a decline in education and

[80] These results are available upon request. Income estimates are slightly smaller when the sample is truncated at the 99th percentile (1.5 percent instead of 2.5 percent).

income in adulthood. No other periods are significant. Furthermore, the estimated relationship between early-life typhoid exposure and adult outcomes is similar to the estimates presented in Table 24. Although not presented in Figure 14, when the outcome is education, the coefficient on typhoid during early life is statistically different from any of the other lifecycle periods (at either the one or five percent level). The estimates for income, however, are less precisely estimated and so the coefficient on early-life exposure is only statistically different from one of the other lifecycle periods.

Figure 14: Average typhoid rates at various stages and adult outcomes



Each regression includes fixed effects for city of birth, year of birth, race, and birth order.

As an additional robustness check, we add state of birth by year of birth fixed effects to equation (3.2). These fixed effects control for any shocks that affected a single cohort at the state level. These shocks could include, compulsory schooling legislation, agricultural shocks and other state-level economic conditions, and exposure to extreme weather conditions. Similar to Table 24, the estimates in Table 26 indicate that the eradication of typhoid fever would have increased educational attainment by 0.09 years (significant at the one percent level) and increased earning by 1.6 percent (significant at the five percent level). We find no evidence that exposure to typhoid during early life affected homeownership, but those that were exposed were less likely to move. Although these results are similar to Table 24, this is not our preferred specification because including state-by-year fixed effects absorbs all of the variation for states that only have one city in the sample.[81] This removes 19 cities from our sample including New Orleans, Baltimore, and Washington DC.

Table 26: OLS results with state-by-year fixed effects

	Years of schooling	ln(income)	Homeowner	Mover
Average typhoid rate	-0.0022***	-0.0004**	-0.0001	-0.0002**
during early life	(0.0006)	(0.0002)	(0.0001)	(0.0001)
Black	-1.7252***	-0.7127***	-0.2425***	0.0284
	(0.1724)	(0.0296)	(0.0196)	(0.0255)
Birth year fixed effects	Y	Y	Y	Y
Birth city fixed effects	Y	Y	Y	Y
Birth order fixed effects	Y	Y	Y	Y
Birth-state-by-birth-year fixed effects	Y	Y	Y	Y
The average effect from	0.0916***	0.0162**	0.0024	0.0101**
eradicating typhoid†	(0.0253)	(0.0079)	(0.0061)	(0.0045)
Observations	184,331	141,857	153,932	189,515
R-squared	0.0551	0.0401	0.0410	0.2391

†The average effect from eradicating typhoid is calculated by multiplying the negative of the coefficient by the average typhoid rate during early life (41.72 deaths per 100,000). Robust standard errors (clustered at the city level) reported in parentheses. * p<.10; ** p<0.05; *** p<0.01

Our final robustness check concerns our birth-city assumption. We assume that all children under the age of 11 in 1900 resided in their city of birth so long as they were born in the same state. This is likely to attenuate the results because children born in the countryside, for instance, but migrated to the city before 1900 will be included in our analysis despite being exposed to a different disease environment. Approximately 6.8 percent of individuals under the age of 11 (that also resided in a city for which we have typhoid data for) were born out of state. Older cohorts are less likely to reside in their city of birth than younger cohorts. To mitigate this problem we replicate Table 24 dropping all individuals born before 1895. These results are

[81] For this reason, one might suggest that we use city specific linear trends. The main results are not robust to the inclusion of city-specific linear trends because our data are at the city-year level and so the inclusion of trends will capture a lot of the variation in typhoid death rates.

presented in Table 27. The results are similar but slightly larger in magnitude, which is consistent with the attenuation argument discussed above. Specifically, we find that the eradication of typhoid fever would have increased education by about two months and increased earnings by about two percent. We also find that those exposed to typhoid were less mobile.

Table 27: OLS results when the sample restricted to those born between 1895 and 1900

	Years of schooling	ln(income)	Homeowner	Mover
Average typhoid rate	-0.0038***	-0.0004*	-0.0002	-0.0004**
during early life	(0.0012)	(0.0003)	(0.0002)	(0.0002)
Black	-1.7519***	-0.7301***	-0.2407***	0.0295
	(0.1834)	(0.0297)	(0.0188)	(0.0261)
Birth year fixed effects	Y	Y	Y	Y
Birth city fixed effects	Y	Y	Y	Y
Birth order fixed effects	Y	Y	Y	Y
The average effect from	0.1576***	0.0184*	0.0069	0.0176**
eradicating typhoid†	(0.0488)	(0.0107)	(0.0101)	(0.0067)
Observations	101,569	79,504	83,754	104,377
R-squared	0.054	0.043	0.036	0.239

†The average effect from eradicating typhoid is calculated by multiplying the negative of the coefficient by the average typhoid rate during early life (41.72 deaths per 100,000). Robust standard errors (clustered at the city level) reported in parentheses. * p<.10; ** p<0.05; *** p<0.01

3.4.c. Semi-parametric results

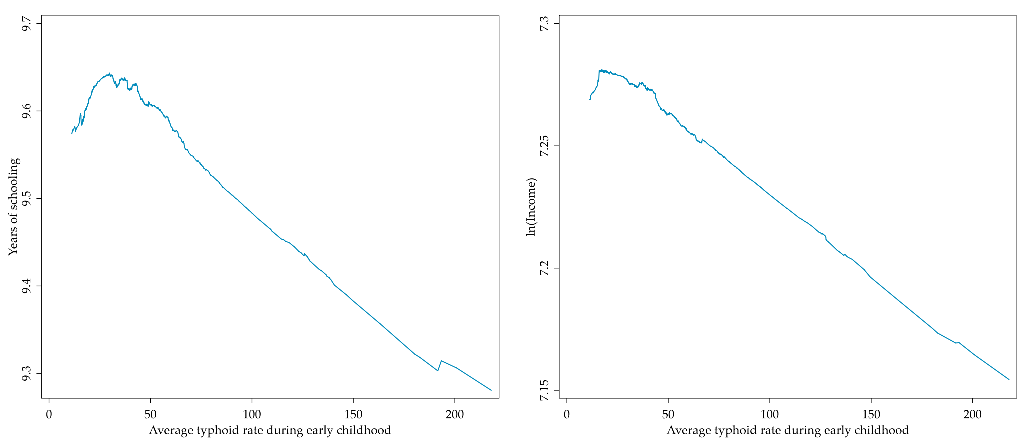
A concern with the analysis above is that it imposes a linear relationship on the data when the data might in fact be related in non-linear ways. To address this concern, we estimate the relationship between typhoid and adult outcomes semi-parametrically. Specifically, we estimate the following equation:

$$y_{jk} = \alpha + f\left(Typhoid_{jk}\right) + \beta\left[black_{jk}\right] + \gamma\left[birth\ order_{jk}\right] + birth\ city\ FE's + birth\ year\ FE's + \epsilon_{jk} \qquad (3.3)$$

this equation is similar to equation (3.2) except that it does not impose a linear relationship between early-life typhoid exposure and adult outcomes. We non-parametrically estimate the relationship between early-life typhoid exposure and adult outcomes using linear partial regression. However, this requires a strict ordering of early-life typhoid rates. We achieve this by collapsing the data at the city-year level.[82] Since we collapse at the city-year level, $black_{jk}$ becomes the percent of the cohort born in city j during year k that is black, and $birth\ order_{jk}$ becomes the average birth order for individuals born in city j during year k.

Figure 15 presents the non-parametric estimates of $f(Typhoid)$. Early-life exposure to typhoid decreases adult earnings and educational attainment above the 10th percentile of the early-life typhoid distribution. Moreover the relationship is approximately linear. Moving from the top of the typhoid distribution to eradication would have increased educational attainment by one-third of a year and increased earnings by about four percent. There does, however, appear to be a positive relationship between zero and 20 deaths per 100,000, but this constitutes less than ten percent of our sample.[83] Overall then, it appears that the linear model adopted in section *3.4.a.* is appropriate.

Figure 15: Semi-parametric estimates of the relationship between typhoid and adult outcomes



This figure presents the semi-parametric estimates from equation (3.2). The average typhoid rate during early life is the average typhoid rate during the year before birth, the year of birth, and the year after birth, and rate is the number of deaths per 100,000.

[82] Linear partial regression requires that we can sort typhoid rates from lowest to highest. If we did not collapse at the city-year level, then there would be many individuals with the same early-life typhoid rates, and the estimates would be sensitive to the sorting order.

[83] This could also be due to survivorship bias (see Bozzoli, Deaton, and Quintana-Domeque, 2009).

3.4.d. Two-stage least squares results

One might be concerned that typhoid during early life is correlated with variables we cannot observer or otherwise directly control for. In particular, two competing hypotheses seem plausible. First, investment in water filtration might be correlated with unobservable investments that also increase human capital. In this case, our OLS estimates would overstate the benefits of water filtration and purification on later life economic outcomes. Second, and alternatively, there was a heavy mortality penalty for living in large, fast growing cities, and this penalty grew larger as city size grew (Cain and Hong 2009; Haines 2001). This suggests that typhoid rates might have been the highest in years of unusually rapid economic growth, and to the extent that we do not fully control for such growth, high typhoid rates might conflate the beneficial health effects of being exposed to economic growth (and high income) early in life with the deleterious effects of impure water. If so, one would want to look at a measure of water quality that captured the effects of only water quality and not economic activity. In the absence of such a measure, our OLS estimates would understate the benefits of water filtration.

To address these concerns, we implement an instrumental variables strategy. This strategy builds on the following logic: because typhoid is a waterborne disease, cities that dump their sewage into a river will increase future typhoid rates for cities downstream. Additionally, the typhoid rates in cities upstream should be exogenous to human capital investments in the receiving city. Of course, as we discussed in Section 3.2.d, typhoid fever was also spread through milk. Accordingly, one might be concerned about overlapping dairy markets as a threat to our exclusion restriction. However during our sample period (1890-1900), diary markets were highly localized, typically concentrated in a sixty-mile radius around the city (Rosenau 1912, pp. 16-20). Because the mean distance between the upstream and downstream cities in our sample is 174 miles, it seems unlikely that upstream milk markets would have somehow influenced downstream typhoid rates. The localization of milk markets was driven by the fact that milk was not transported in refrigerated rail cars during the 1890s and early 1900s (United States 1912).

Eighteen of the 75 cities used in the previous analysis lie downstream from another city for which we have typhoid data. We confirm flow direction for each river using data from the

United States Geological Survey.[84] Cities that are upstream (the feeder cities) dump their sewage into the river. This increases the typhoid rates in cities downstream (the receiving city). Thus, we use typhoid rates in the feeder city an instrument for typhoid rates in the receiving city. Whether we should use contemporaneous typhoid rates or the rates lagged by one year depends on the distance between the two cities and the flow rate of the river. We find similar results regardless of whether we use the contemporaneous or lagged typhoid rate, but lagged typhoid rates produce a stronger first stage.

Because only a subset of our initial sample lies downstream from another city for which we have typhoid data, it is perhaps useful to show that our main results hold for this subset of cities. Accordingly, we replicate our main analysis (restricting to the set of downstream cities) in Table 28.[85] The results presented in Table 28 are slightly larger but qualitatively similar to the main results presented in Table 24. Specifically, we find that eradication of typhoid fever would have increased educational attainment by 1.4 months instead of 1.1 and would have increased adult earnings by about 1.7 percent instead of 1.3.

[84] Specifically, we identified flow direction using the USGS/National Map Streamer tool. This information is available as a web app at: http://nationalmap.gov/streamer/webApp/welcome.html

[85] Because we only have 18 cities in this sample, we use robust standard errors instead of clustered standard errors in Table 8.

Table 28: The relationship between typhoid and adult outcomes (IV sample only)

	Years of schooling	ln(income)	Homeowner	Mover
Average typhoid rate during early life	-0.0028***	-0.0004*	0.0001	-0.0002
	(0.0009)	(0.0002)	(0.0001)	(0.0002)
Black	-1.6646***	-0.6918***	-0.2135***	0.0299
	(0.2545)	(0.0483)	(0.0192)	(0.0495)
Birth year fixed effects	Y	Y	Y	Y
Birth city fixed effects	Y	Y	Y	Y
Birth order fixed effects	Y	Y	Y	Y
The average effect from eradicating typhoid†	0.1159***	0.0173*	-0.0052	0.0073
	(0.0370)	(0.0086)	(0.0058)	(0.0078)
Observations	73,496	56,254	61,398	76,085
R-squared	0.0511	0.0427	0.0329	0.0902

†The average effect from eradicating typhoid is calculated by multiplying the negative of the coefficient by the average typhoid rate during early life (41.72 deaths per 100,000). Robust standard errors reported in parentheses.
* p<.10; ** p<0.05; *** p<0.01

Table 29 presents our results using lagged typhoid rates in the feeder city as an instrument for typhoid rates in the receiving city. Lagged typhoid rates in the feeder city are a strong predictor of typhoid rates in the receiving city; an additional 100 deaths per 100,000 in the feeder city increases the typhoid death rate in the receiving city by 8 in the following year. The F-statistics associated with this estimate range from 517.81 to 671.12 and therefore suggest that lagged typhoid rates from the feeder city are a strong instrument. In the second stage we find that typhoid rates during early life decrease educational attainment and earnings, although only the first estimate is statistically significant at the five percent level.[86] The estimate on earnings, while imprecisely estimated, has the same sign and much larger coefficient in absolute value than the OLS result presented in Table 24. These results indicate that the eradication of typhoid would have increased schooling by nine months, and would have increased income by 9.8 percent.

[86] Although the results reported in Table 5 include birth-order and race fixed effects, omitting these controls does not affect the results. Results without these controls available upon request.

Table 29 also indicates that high typhoid rates during early life reduced mobility, i.e. the likelihood that the individual would reside in their birth city as an adult. The results here are considerably larger in magnitude than those obtained with OLS, and using a Durbin-Wu-Hausmann test we can reject the consistency between the IV and OLS estimates at the 10 percent level. (We cannot reject at the 5-percent significance level, however.) This suggests that the difference in magnitudes stems from the fact we not able to fully control for the effects of positive economic growth using OLS.

Table 29: 2SLS estimates of early-life typhoid on adult outcomes

	Years of schooling	ln(income)	Homeowner	Mover
Instrumented typhoid rate	-0.0189**	-0.0023	0.0014	-0.0025*
	(0.0091)	(0.0025)	(0.0016)	(0.0013)
Average effect from eradicating typhoid†	0.7869**	0.0976	-0.0578	0.1037*
	(0.3802)	(0.1035)	(0.0655)	(0.0562)
Hausman test (p-value)	0.0755	0.4344	0.4187	0.0839
Observations	73496	56254	61398	76085
R-squared	0.048	0.042	0.032	0.088
First stage				
Lagged typhoid rate in feeder city	0.0872***	0.0921***	0.0883***	0.0881***
	(0.0035)	(0.0040)	(0.0038)	(0.0034)
f-statistic	628.124	517.813	542.895	671.12
Observations	73,496	56,254	61,398	76,085
R-squared	0.757	0.757	0.759	0.755

† The average effect from eradicating typhoid is calculated by multiplying the negative of the coefficient by the average typhoid rate during early life (41.72 deaths per 100,000). Robust standard errors reported in parenthesis. Each regression includes fixed effects for city of birth, year of birth, race, and birth order.
* p<.10; ** p<0.05; *** p<0.01

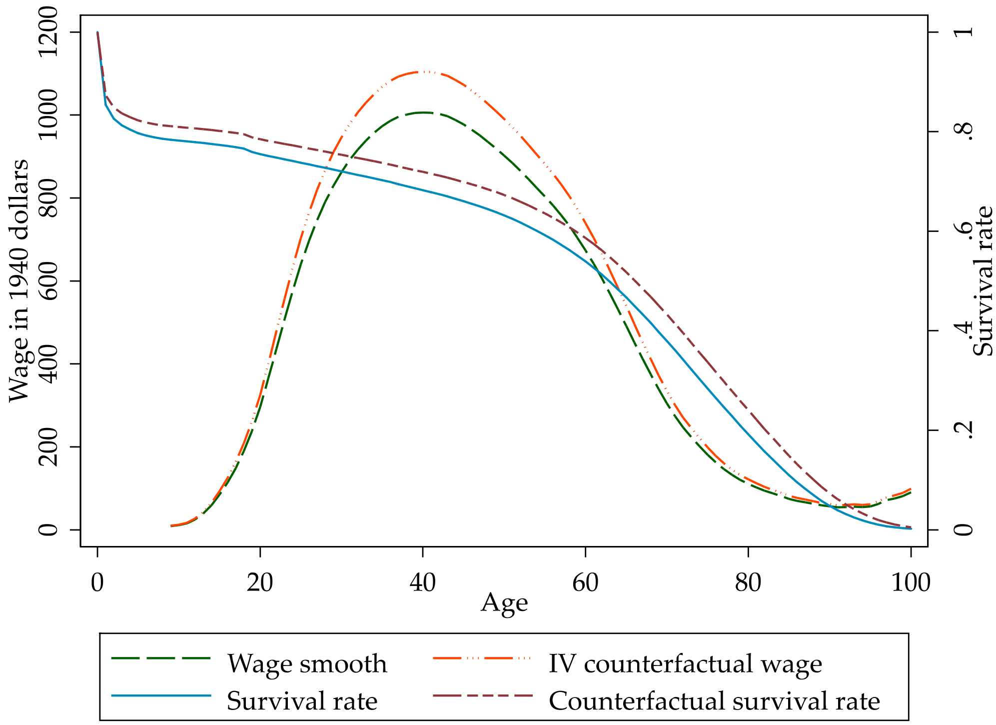
3.5 COST-BENEFIT ANALYSIS

The previous section illustrates that eradicating typhoid fever would have increased educational attainment by one to nine months and income by one to nine percent. These estimates raise the question of whether the net present value of the increase in wages was enough to offset the costs associated with eradicating typhoid fever. Cutler and Miller (2005) have analyzed the benefits of adopting water purification technologies using the value of a statistical life and find that the benefits outweigh the costs by a ratio of 23 to 1. In our analysis, we ignore the gains from additional life years and instead focus on whether the discounted increase in earnings would have been sufficient to cover the costs of eradicating typhoid fever.

To analyze the benefits from typhoid eradication, we need the following information: the probability that an individual survives to a given age with and without the intervention, average income by age with and without the intervention, and the number of individuals in a cohort who would benefit from the water infrastructure. To analyze the costs, we need to know the total costs of municipal water systems and how frequently these systems need to be replaced. Lastly, to compare the net present values, we need real interest rates.

The survival probability, S_a, is the probability that an individual survives to age a. We use the survival probabilities for males born in 1900 from the Social Security life tables. Cutler and Miller (2005) find that mortality fell by 13% after the introduction of clean water technologies. Accordingly, we adjust the 1900 survival probabilities to reflect this change. We use the 1940 census to obtain the wage profile for males. Specifically, we obtain average earnings by age using a local polynomial smooth for all males. For the counterfactual wage distribution, we scale these averages by one to nine percent, which corresponds to the OLS and IV estimates reported in Tables 4 and 7. Figure 16 plots the baseline and counterfactual survival probabilities as well as the baseline and counterfactual wages. Eradicating typhoid fever has two effects on wages. First, it increases the average wage. Second, it increases the probability that an individual will survive until that age.

Figure 16: Earnings profile and survival curve



Survival rate obtained from Social Security Administration life tables for the 1900 male birth cohort. Wage obtained from IPUMS one percent sample of males in 1940. The counterfactual survival rate is adjusted using the estimated 13% decline in mortality rates reported in Cutler and Miller (2005). The IV counterfactual wage is adjusted using the 9% estimate from Table 29.

We assume that the average cohort of males born in a city is 20,000, which is approximately the number of males born in Chicago in 1900.[87] Finally, to obtain the costs we use the numbers reported in Cutler and Miller (2005), which assumes that the average cost of the waterworks for a large city was 22.8 million dollars in 1940 and that the waterworks must be replaced every ten years.

The previous assumptions underestimate the gains from eradicating typhoid fever. First, we assume that female earnings were unaffected by typhoid. Second, we assume that the only benefit from reduced mortality was increasing the probability of receiving future earnings. Third, we assume that the entire waterworks must be replaced every ten years, when in reality many parts are likely to function for longer. Furthermore, we assume that the construction of the

[87] We obtain 20,000 by taking the number of males in Chicago that were age 0 in the IPUMS 5% sample of the 1900 census and multiplying it by 20.

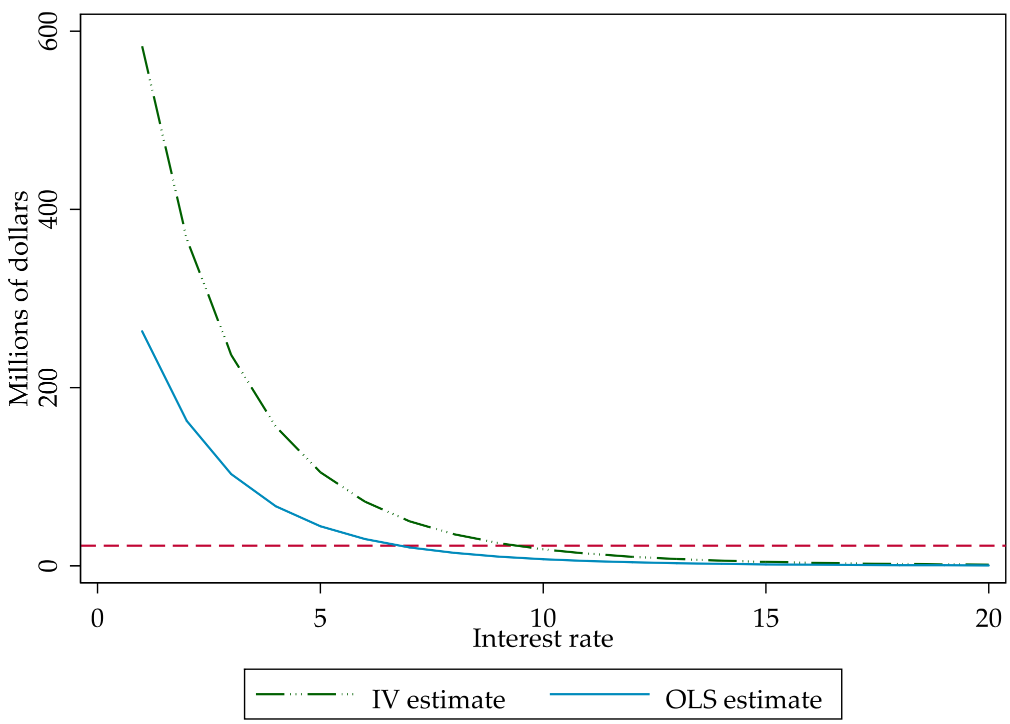
waterworks was necessary to eradicate typhoid fever, but one could argue that the marginal cost of chlorinating or filtering water was sufficient to eradicate typhoid fever.

We calculate the benefits to eradicating typhoid fever using equation three, where ${S_a}'$ is the counterfactual survival probability and ${W_a}'$ is the counterfactual wage. The waterworks lasts T years, N is the cohort size, and r is the real interest rate.

$$NPV = \sum_{i=0}^{T} \frac{N}{(1+r)^i} * \sum_{a=0}^{101} \frac{({S_a}'{W_a}') - (S_a W_a)}{(1+r)^a} \tag{3.4}$$

Figure 17 graphs these benefits for various interest rates for both the OLS and IV counterfactual wages. The horizontal line corresponds to the cost of eradicating typhoid fever, 22.8 million in 1940 dollars. Figure 17 shows that for our OLS estimates and any real interest rate under seven percent, the increase in earnings alone was sufficient to offset the cost of eradicating typhoid fever. For our IV estimates, the break-even real interest rate increases to ten percent.

Figure 17: Net present value of typhoid eradication



The net present value of the benefits is obtained from equation (3.3) for various interest rates. The horizontal line corresponds to the estimated cost of the waterworks, 22.8 million dollars in 1940 (obtained from Cutler and Miller, 2005).

3.6 DISCUSSION AND CONCLUSION

Between 1900 and 1940 mortality in the United States fell by nearly 40 percent. Approximately half of this decline was the result of investment in water purification technologies and the eradication of waterborne diseases such as typhoid fever. There have been a number of previous studies estimating the social rate of return to water purification measures, but all of these studies focus on the gains associated with reductions in mortality (e.g., Cutler Miller 2005; Ferrie and Troesken 2008). Yet because typhoid was such a virulent disease and had such a low case fatality rate, there is good reason to believe that its effects on morbidity and long-term human capital formation were substantial. Accordingly, in this paper, we explore how eliminating early-life exposure to typhoid fever affected economic outcomes in later life. Our laboratory consists of urban residents in large American cities during the late-nineteenth and early twentieth century.

In our analysis, we explore how early life exposure to typhoid fever (our primary indicatory of water quality) influenced later life outcomes in terms of income, educational attainment, home ownership, and geographic mobility. Using parametric, semi-parametric, and IV approaches, our results indicate that the eradication of typhoid fever, which cities achieved by adopting clean water technologies, would have increased educational attainment by one to nine months and earnings would have increased by between one and nine percent. A simple cost-benefit analysis reveals that the increase in earnings from eradicating typhoid fever was more than sufficient to offset the costs of eradication. When one considers that our calculations ignore the changes in mortality captured by Cutler and Miller (2005) and other researchers, the evidence that investments in water purification have very high rates of social return seems unassailable. These results have important policy implications for developing countries that have yet to adopt water purification technologies.

BIBLIOGRAPHY

Alesina, Alberto. 1988. "Credibility and policy convergence in a two-party system with rational voters." *The American Economic Review* 78 (4): 796-805.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public goods and ethnic divisions." *The Quarterly Journal of Economics* 114 (4): 1243-84.

Alesina, Alberto, and Eliana La Ferrara. 2005. "Ethnic diversity and economic performance." *Journal of Economic Literature* 43 (3): 762-800.

Alesina, Alberto, and Ekaterina Zhuravskaya. 2011. "Segregation and the quality of government in a cross section of countries." *The American Economic Review* 101 (5): 1872-1911.

Almond, Douglas and Janet Currie. 2011. "Human capital development before age five." *Handbook of Labor Economics,* 4: 1315-1486.

Almond, Douglas, Janet Currie, and Mariesa Herrmann. 2012. "From infant to mother: Early disease environment and future maternal health." *Labour Economics* 19 (4): 475-83.

Alquist, Ron and Benjamin Chabot. 2011. "Did gold-standard adherence reduce sovereign capital costs?." *Journal of Monetary Economics* 58 (3): 262-72.

American Public Health Association. Report of Committee on Milk Supply. *Pasteurization of milk*. Boston, American Public Health Association, 1920.

Beemer, Jeffrey, Douglas Anderton, and Susan Hautaniemi Leonard. 2005. "Sewers in the city: A case study of individual-level mortality and public health initiatives in Northampton, Massachusetts, at the turn of the century." *Journal of the History of Medicine and Allied Sciences* 60 (1): 42-72.

Ben-Ner, Avner, John-Gabriel Licht and Jin Park. 2014. "Empirical evidence on diversity and performance in teams: The roles of task focus, status and tenure." Working Paper.

Benjamin, David, and Mark Wright. 2009. "Recovery before redemption: A theory of delays in sovereign debt renegotiations." *Mimeo, University of California at Los Angeles*.

Berkowitz, Daniel, and Karen Clay. 2005. "American civil law origins: Implications for state constitutions." *American Law and Economics Review* 7 (1): 62-84.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?." *Quarterly Journal of Economics* 119 (1): 249-75.

Besley, Timothy, and Stephen Coate. 2003. "Centralized versus decentralized provision of local public goods: a political economy approach." *Journal of Public Economics* 87 (12): 2611-37.

Besley, Timothy, Torsten Persson, and Daniel Sturm. 2010. "Political competition, policy and growth: theory and evidence from the US." *The Review of Economic Studies* 77 (4): 1329-52.

Bhalotra, Sonia, Irma Clots-Figueras, Guilhem Cassan, and Lakshmi Iyer. 2013. "Religion, politician identity and development outcomes: Evidence from India." *Journal of Economic Behavior & Organization* 104: 4-17.

Bogart, Ernest. 1923. *Internal improvements and state debt in Ohio*. Longmans, Green & Co.

Bordo, Michael and Hugh Rockoff. 1996. "The gold standard as a “good housekeeping seal of approval”." *The Journal of Economic History* 56 (2): 389-428.

Boustan, Leah, Fernando Ferreira, Hernan Winkler, and Eric Zolt. 2013. "The effect of rising income inequality on taxation and public expenditures: Evidence from US municipalities and school districts, 1970–2000." *Review of Economics and Statistics* 95 (4): 1291-1302.

Bozzoli, Carlos, Angus Deaton, and Climent Quintana-Domeque. 2009. "Adult height and childhood disease." *Demography* 46 (4): 647-69.

Budd, William. 1918. “Typhoid fever: Its nature, mode of spreading, and prevention.” Reprinted in *American Journal of Public Health* 8 (8)*:* 61-612.

Bulletin of the Chicago School of Sanitary Instruction. (various issues)

Burnham, W. Dean. 1985. “Partisan division of American state governments, 1834-1985”. Conducted by Massachusetts Institute of Technology. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], doi:10.3886/ICPSR00016.v1.

Cain, Louis. 1977. *Sanitation strategy for a lakefront metropolis*, DeKalb: Northern Illinois University Press.

Cain, Louis, and Sok Chul Hong. 2009. "Survival in 19th century cities: The larger the city, the smaller your chances," *Explorations in Economic History*, 46 (4): 450-63.

Campbell, John, Andrew Lo, and A. Craig MacKinlay. 1997. *The econometrics of financial markets* 2nd edition.

Cascio, Elizabeth and Ebonya Washington. 2014. "Valuing the vote: The redistribution of voting rights and state funds following the voting rights act of 1965." *The Quarterly Journal of Economics* 129 (1): 379-433.

Case, Anne, and Christina Paxson. 2009. "Early life health and cognitive function in old age." *The American Economic Review* 99 (2): 104-9.

Caughey, Devin, and Jasjeet Sekhon. 2011. "Elections and the regression discontinuity design: Lessons from close us house races, 1942–2008." *Political Analysis* 19 (4): 385-408.

Chicago Bureau of Public Efficiency. 1917. *The water works system of the city of Chicago*. No publisher listed.

Chicago Daily Inter-Ocean (newspaper, various issues)

Chicago Department of Health. 1907. *Report of the department of health of the city of Chicago.*

Clark, Gregory. 2008. *A farewell to alms: A brief economic history of the world.* Princeton University Press.

Clark, Gregory. 1996. "The political foundations of modern economic growth: England, 1540-1800." *Journal of Interdisciplinary History* 26 (4): 563-87.

Costa, Dora. 2000. "Understanding the twentieth-century decline in chronic conditions among older men." *Demography* 37 (1): 53-72.

Cox, Gary. 2012. "Was the Glorious Revolution a constitutional watershed?." *Journal of Economic History* 72 (3): 567-600.

Cruces, Juan and Christoph Trebesch. 2013. "Sovereign defaults: The price of haircuts." *American Economic Journal: Macroeconomics* 5 (3): 85-117.

Currie, Janet, Joshua Graff Zivin, Katherine Meckel, Matthew Neidell, and Wolfram Schlenker. 2013. "Something in the water: contaminated drinking water and infant health." *Canadian Journal of Economics* 46 (3): 791-810.

Curschmann, Heinrich, and Alfred Stengel. 1902. *Typhoid fever and typhus fever*. W.B. Saunders.

Cutler, David, and Grant Miller. 2005. "The role of public health improvements in health advances: the twentieth-century United States." *Demography* 42 (1): 1-22.

Dahlberg, Matz, Karin Edmark, and Heléne Lundqvist. 2012. "Ethnic diversity and preferences for redistribution." *Journal of Political Economy* 120 (1): 41-76.

Dove, John. 2012. "Credible commitments and constitutional constraints: state debt repudiation and default in nineteenth century America." *Constitutional Political Economy* 23 (1): 66-93.

Dublin, Louis. 1915. "Typhoid fever and its sequelae." *American Journal of Public Health* 5 (1): 20-7.

Eggers, Andrew, Anthony Fowler, Jens Hainmueller, Andrew Hall, and James Snyder. 2015. "On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races." *American Journal of Political Science* 59 (1): 259-74.

Ellms, Joseph. 1913. "Disinfection of public water supplies: Why the purification of drinking water should be supplemented by disinfection—the uses of chlorine, ozone, and ultra-violet light." *The American City*, 22: 564-68.

Engineering News, May, 1913, p. 1087

English, William. 1996. "Understanding the costs of sovereign default: American state debts in the 1840's." *The American Economic Review* 86 (1): 259-75.

Ferreira, Fernando, and Joseph Gyourko. 2009. "Do political parties matter? Evidence from US cities." *The Quarterly Journal of Economics* 124 (1): 399-422.

Ferreira, Fernando and Joseph Gyourko. 2014. "Does gender matter for political leadership? The case of US mayors." *Journal of Public Economics* 112: 24-39.

Ferrie, Joseph and Werner Troesken. 2005. *Death and the city: Chicago's mortality transition, 1850-1925*. No. w11427. National Bureau of Economic Research.

Ferrie, Joseph and Werner Troesken. 2008. "Water and Chicago's mortality transition, 1850–1925." *Explorations in Economic History* 45 (1): 1-16.

Foley-Fisher, Nathan, and Eoin McLaughlin. 2014. "State dissolution, sovereign debt and default: Lessons from the UK and Ireland, 1920-1938." QUCEH Working Paper Series.

Frost, W.H. 1916. "Relationship of milk Supplies to typhoid fever," *Public Health Reports*, 31 (48): 3291-3302.

Fuertes, James. 1897. *Water and Public Health: The Relative Purity of Waters from Different Sources*. J. Wiley & sons

Gisselquist, Rachel. 2013. "Ethnic divisions and public goods provision, revisited." *Ethnic and Racial Studies* 37 (9): 1-23.

Grimmer, Justin, Eitan Hersh, Brian Feinstein, and Daniel Carpenter. 2011. "Are close elections random?." *Unpublished manuscript*.

Haines, Michael. 2001. "The urban mortality transition in the United States, 1800-1940." *Annales de Démographie Historique, 101: 33-64*

Hicks, Henry and Herbert French. 1905. "Typhoid fever and pregnancy, with special reference to fetal infection." *The Lancet* 165 (4266): 1491-3.

Hopkins, Daniel. 2011. "The limited local impacts of ethnic and racial diversity." *American Politics Research* 39 (2): 344-79.

Kesztenbaum, Lionel and Jean-Laurent Rosenthal. 2007. "Income versus sanitation; mortality decline in Paris, 1880-1914." *Perspectives in Biology and Medicine* 50 (4): 585-602.

Kettell, Thomas. 1851. "Constitutional reform" *The United States Magazine and Historical Review* 29 (157).

Kothari, S. P., and Jerold Warner. 2006. "Econometrics of event studies." *Handbook of Empirical Corporate Finance* 1: 4-32.

La Ferrara, Eliana. 2002. "Self- help groups and income generation in the informal settlements of Nairobi." *Journal of African Economies* 11 (1): 61-89.

La Ferrara, Eliana, and Angelo Mele. 2006. "Racial segregation and public school expenditure." *Unpublished manuscript*.

Lee, David, Enrico Moretti, and Matthew Butler. 2004. "Do voters affect or elect policies? Evidence from the US House." *The Quarterly Journal of Economics* 119 (3): 807-59.

Levitt, Steven. 1996. "How do senators vote? Disentangling the role of voter preferences, party affiliation, and senator ideology." *The American Economic Review* 86 (3): 425-41.

Mauro, Paolo, Nathan Sussman, and Yishay Yafeh. 2002. "Emerging market spreads: then versus now." *Quarterly Journal of Economics* 117 (2): 695-733.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142 (2): 698-714.

McLeod, Poppy, Sharon Lobel, and Taylor Cox. 1996. "Ethnic diversity and creativity in small groups." *Small group research* 27 (2): 248-64.

Medical News (various issues)

Melosi, Martin. 2000. *The sanitary city: Environmental services in urban America from colonial times to the present*. Johns Hopkins University Press.

Mitchener, Kris and Marc Weidenmier. 2005. "Empire, public goods, and the Roosevelt corollary." *Journal of Economic History* 65 (3): 658-92.

Mitchener, Kris and Marc Weidenmier. 2010. "Supersanctions and sovereign debt repayment." *Journal of International Money and Finance* 29 (1): 19-36.

Municipal and County Engineering (various issues)

Neal, Larry. 1993. *The rise of financial capitalism: International capital markets in the age of reason*. Cambridge University Press.

Neal, Larry. 1992. "The disintegration and re-integration of international capital markets in the 19th century." *Business and Economic History*, 21 (2): 84-96.

North, Douglass and Barry Weingast. 1989. "Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England." *Journal of Economic History* 49 (4): 803-32.

Oates, Wallace. 1972. *Fiscal federalism*. New York: Harcourt Brace Jovanovich.

Obstfeld, Maurice and Alan Taylor. 2003. "Sovereign risk, credibility and the gold standard: 1870–1913 versus 1925–31." *The Economic Journal* 113 (487): 241-75.

Ottaviano, Gianmarco and Giovanni Peri. 2005. "Cities and cultures." *Journal of Urban Economics* 58 (2): 304-37.

Ottaviano, Gianmarco and Giovanni Peri. 2006. "The economic value of cultural diversity: evidence from US cities." *Journal of Economic Geography* 6 (1): 9-44.

Pande, Rohini. 2003. "Can mandated political representation increase policy influence for disadvantaged minorities? Theory and evidence from India." *The American Economic Review* 93 (4): 1132-51.

Ratchford, Benjamin. 1966. *American state debts*. AMS Press.

Reardon, Sean and Glenn Firebaugh. 2002. "Measures of multigroup segregation." *Sociological methodology* 32 (1): 33-67.

Reynal-Querol, Marta. 2002. "Ethnicity, political systems, and civil wars." *Journal of Conflict Resolution*, 46 (1): 29-54.

Riggall, F., G. Salkind, and W. Spizllacy. 1974. "Typhoid fever complicating pregnancy." *Obstetrics & Gynecology* 44 (1): 117-121.

Rosenau, M.J. 1912. *The Milk Question*. Boston and New York: Houghton Mifflin Company.

Rugh, Jacob and Jessica Trounstine. 2011. "The provision of local public goods in diverse communities: Analyzing municipal bond elections." *The Journal of Politics* 73 (4): 1038-50.

Saiegh, Sebastian. 2013. "Political institutions and sovereign borrowing: evidence from nineteenth-century Argentina." *Public Choice* 156 (1): 1-15.

Sedgwick, William. 1902. *Principles of sanitary science and the public health: with special reference to the causation and prevention of infectious diseases*. Macmillan Company; London, Macmillan and Company, Limited.

Sedgwick, William and J. Scott MacNutt. 1910. "On the Mills-Reincke phenomenon and Hazen's theorem concerning the decrease in mortality from diseases other than typhoid fever following the purification of public water-supplies." *Journal of Infectious Diseases* 7 (4): 489-564.

Seoud, Muhieddine, George Saade, Marwan Uwaydah, and Ramez Azoury. 1988. "Typhoid fever in pregnancy." Obstetrics & Gynecology 71 (5): 711-4.

Shambaugh, Jay, Ricardo Reis, and Helene Rey. 2012. "The euro's three crises [with comments and discussion]." *Brookings Papers on Economic Activity*: 157-231.

Shore, Lynn, Beth Chung-Herrera, Michelle Dean, Karen Holcombe Ehrhart, Don Jung, Amy Randel, and Gangaram Singh. 2009. "Diversity in organizations: where are we now and where are we going?." *Human Resource Management Review* 19 (2): 117-33.

Sinha, Anju, Sunil Sazawal, Ramesh Kumar, Seema Sood, Vankadara P. Reddaiah, Bir Singh, Malla Rao, Abdolla Naficy, John D. Clemens, and Maharaj K. Bhan. 1999. "Typhoid fever in children aged less than 5 years." *The Lancet* 354 (9180): 734-7.

Stasavage, David. 2002. "Credible commitment in early modern Europe: North and Weingast revisited." *Journal of Law, Economics, and Organization* 18 (1): 155-86.

Stasavage, David. 2007. "Partisan politics and public debt: The importance of the 'Whig Supremacy' for Britain's financial revolution." *European Review of Economic History* 11 (1): 123-53.

Stasavage, David. 2008. "Cities, constitutions, and sovereign borrowing in Europe, 1274-1785." *International Organization* 61 (3): 489-525.

Stevenson, Charles S., Anthony J. Glazko, E. Clark Gillespie, and John B. Maunder. 1951. "Treatment of typhoid in pregnancy with chloramphenicol." *Journal of the American Medical Association* 146 (13): 1190-2.

Sulaiman, K., and A. R. Sarwari. 2007. "Culture-confirmed typhoid fever and pregnancy." *International Journal of Infectious Diseases* 11 (4): 337-41.

Summerhill, William. 2006. "Sovereign commitment and financial underdevelopment in imperial Brazil." *Mimeo*.

Sussman, Nathan and Yishay Yafeh. 2000. "Institutions, reforms, and country risk: lessons from Japanese government debt in the Meiji era." *Journal of Economic History* 60 (2): 442-67.

Sussman, Nathan and Yishay Yafeh. 2006. "Institutional reforms, financial development and sovereign debt: Britain 1690-1790." *Journal of Economic History* 66 (4): 906-35.

Sylla, Richard, Jack Wilson, and Robert Wright. 2002. "Price quotations in early U.S. securities markets, 1790–1860." http://www.eh.net/databases/early-us-securities-prices.

Sylla, Richard, Jack Wilson, and Robert Wright. 2006. "Integration of trans-Atlantic capital markets, 1790–1845." *Review of Finance* 10 (4): 613-44.

Thorpe, Francis. 1909. *The federal and state constitutions: Colonial charters, and other organic laws of the states, territories, and colonies now or heretofore forming the United States of America*. US Government Printing Office.

Tomz, Michael and Mark Wright. 2007. "Do countries default in 'bad times'?." *Journal of the European Economic Association* 5 (2- 3): 352-60.

Tomz, Michael and Mark Wright. 2013. "Empirical research on sovereign debt and default." *Annual Review of Economics* 5 (1): 247-72.

Troesken, Werner. *Water, race, and disease*. MIT Press, 2004.

United States Bureau of the Census. 1915. *General Statistics of Cities*. Washington, DC; U.S. Government printing office.

United States Public Health and Marine Hospital Service. 1909. *Milk and its relation the public health. Hygienic Laboratory*---Bulletin No. 56. Washington, D.C.: Government Printing Office, 1912.

Vogl, Tom. 2014. "Race and the politics of close elections." *Journal of Public Economics* 109: 101-13.

Wallis, John. NBER/University of Maryland State Constitution Project, http://www.stateconstitutions.umd.edu

Wallis, John. 2000. "American government finance in the long run: 1790 to 1990." *The Journal of Economic Perspectives* 14 (1): 61-82.

Wallis, John. 2005. "Constitutions, corporations, and corruption: American states and constitutional change, 1842 to 1852." *Journal of Economic History* 65 (1): 211-56.

Wallis, John, Richard Sylla, and Arthur Grinath III. 2004. "Sovereign debt and repudiation: The emerging-market debt crisis in the US States, 1839-1843". *National Bureau of Economic Research*, no. w10753.

Washington, Ebonya. 2008. "Female socialization: How daughters affect their legislator fathers' voting on women's issues." *The American Economic Review* 98 (1): 311-32.

Weingast, Barry, Kenneth Shepsle, and Christopher Johnsen. 1981. "The political economy of benefits and costs: A neoclassical approach to distributive politics." *The Journal of Political Economy* 89 (4): 642-64.

Wells, John and Douglas Wills. 2000. "Revolution, restoration, and debt repudiation: The Jacobite threat to England's institutions and economic growth." *Journal of Economic History* 60 (2): 418-41.

Whipple, George. 1908. *Typhoid fever; its causation, transmission and prevention*. J. Wiley & sons.

Wright, Robert. 2002. *The Wealth of Nations Rediscovered: Integration and Expansion in American Financial Markets, 1780-1850*. Cambridge University Press.