Essays in Political Economy

by

Camilo García-Jimeno
M.A., Universidad de los Andes (2005)
B.A. Universidad de los Andes (2004)

Submitted to the Department of Economics
in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 2011

© 2011 Camilo García-Jimeno. All Rights Reserved

Signature of Author

Department of Economics
May 15, 2011

Certified by
Daron Acemoglu
Elizabeth and James Killian Professor of Economics, MIT
Thesis Advisor

Certified by
James A. Robinson
David Florence Professor of Government, Harvard University
Thesis Advisor

Accepted by
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics, MIT
Chairman, Departmental Committee on Graduate Studies
Table of Contents

Abstract.........................................................................................................................p.5


Chapter 2: Entrenchment (and Encroachment) Dynamics........................................p. 93

Chapter 3: The Myth of the Frontier.................................................................p.165
Essays in Political Economy

by

Camilo García-Jimeno

Submitted to the Department of Economics on May 15th, 2011 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Economics

ABSTRACT

This dissertation consists of three essays. The first chapter is an empirical investigation of social change, looking at the Prohibition Era in the U.S. It explores how the implementation of policies affects the evolution of beliefs about their effects, giving rise to a feedback between preferences and policy choices. Using city-level data on law enforcement and crime, it estimates a structural model where crime outcomes are the result of Prohibition enforcement, and lead to changes in public opinion about Alcohol-related policies. Enforcement depends on moral views and beliefs, but only beliefs are shaped by the outcomes of past policies. The model can account for the variation in public opinion changes, and for the heterogeneous responses of enforcement and violence across cities. Its estimates are used to perform a series of counterfactual exercises.

The second chapter is a theoretical investigation of entrenchment and encroachment of rulers. It studies the strategic interaction between competition and ratchet effect incentives in a coalition-formation game of incomplete information. Rulers require the support of a subset of politically powerful groups to remain in power. These have private information about their cost of providing political support. A ruler can attempt to exploit the competitive nature of the coalition formation game to induce revelation. Its ability to do so determines the extent of entrenchment and encroachment. By restricting attention to Markov Perfect Bayesian equilibria, the model shows that limited learning is possible, and that learning dynamics are shaped by an informational commitment problem arising when rulers are “too optimistic”.

In joint work with James Robinson, the final chapter is a comparative empirical study of the impact of Frontier availability on long-run development across the Americas. It calls into question the notion of American exceptionalism due to its Westward Frontier, first proposed by Frederick J. Turner. Almost every country in the Americas had a substantial Frontier, but its allocation varied due to differences in the quality of political institutions around the mid-19th century, making the effect of the Frontier conditional on political institutions at the time of Frontier expansion. The empirical evidence is consistent with this “conditional Turner thesis”.

Thesis Supervisor: Daron Acemoglu
Title: Elizabeth and James Killian Professor of Economics
Chapter 1: The Political Economy of Moral Conflict: An Empirical Study of Learning and Law Enforcement under Prohibition*

Camilo García-Jimeno.†

June, 2011

Abstract

The U.S. Prohibition experience shows a remarkable policy reversal. In only 14 years, a drastic shift in public opinion necessitated two amendments of the U.S. Constitution. The adoption of many other policies and laws is similarly driven by initially optimistic beliefs about potential costs of their enforcement. Their implementation, in turn, affects the evolution of beliefs, giving rise to an endogenous feedback between preferences and policy choices. This paper uses data on U.S. cities during the Prohibition Era to investigate how changes in beliefs about the enforcement costs of Prohibition affected the mapping from moral views to policy outcomes, ultimately resulting in the repeal of Constitutional Prohibition. It first develops a dynamic equilibrium model in which communities make collective choices about law enforcement. Individuals differ in their baseline moral views about alcohol consumption and in their priors about the effects of Prohibition on crime. While both beliefs and moral views determine policy outcomes through the process of democratic decision-making, beliefs are in turn shaped by the outcomes of past policies. The model is estimated using a maximum likelihood approach on city-level data on public opinion, police enforcement, crime, and alcohol-related legislation. The estimated model can account for the variation in public opinion changes, and for the heterogeneous responses of enforcement and violence across cities. Shutting down the learning channel significantly limits the model's ability to match the moments of interest. The paper concludes with a series of counterfactual exercises that explore the equilibrium implications of changes in moral views, priors concerning the costs of enforcement, the degree of polarization in society, and the local political environment.

*This paper constitutes the first chapter of my Ph.D. dissertation at MIT. I am extremely grateful to Daron Acemoglu, James Robinson, and James Snyder for their advice and guidance, and to Abhijit Banerjee and the participants at MIT’s Political Economy Breakfast for their helpful suggestions. I also thank Maria Angelica Bautista, Angela Fonseca, Juanita Gonzalez, Catalina Herrera, Lucas Higuera, Thomas Morin, Michael Peters, and Maria Fernanda Rosales for their help at different stages of the project, and the staff at Harvard’s Law Library Special Collections Room for their kind help. I also gratefully acknowledge the financial support of MIT’s George and Obie Schultz Fund, and of the Banco de la Republica de Colombia. Errors are all my own.

†MIT, Department of Economics. 50 Memorial Drive, Of. E52-201, Cambridge, MA 02139. cgarcia@mit.edu.
1 Introduction

In Individual Choice and Social Values, Arrow (1963) argues that a proper understanding of collective choices requires taking into account the moral views of individuals because, as part of their preferences, they are analytically similar to externalities1. This insight proves particularly relevant in contemporary societies, where cultural heterogeneity is widespread and has been increasing over time, and polities are constituted by peoples with varying cultural backgrounds, and thus, different moral views. Indeed, differences in moral views have become a mayor source of disagreement about policy issues in many Western societies 2.

How differences in moral views affect policies is inexorably linked to individuals’ beliefs about the implications of bans on certain activities, practices and expressions. While moral views and beliefs are mutually self-reinforcing, for example because those who find certain behaviors abhorrent also think that banning them can be effective and would have only minor unintended consequences, there is also a fundamental difference between moral views and beliefs. Moral views are slow-changing or even fixed, whereas beliefs about the implications of different types of bans and restrictions are frequently subject to a large extent of uncertainty, and can change rapidly as individuals observe their outcomes over time. Indeed, learning may be one reason why societies sometimes undergo radical social change and policy reform away from policies originally motivated by moral views, such as during the U.S. alcohol Prohibition experience of the early 20th Century. In this paper I argue that the reversal of Prohibition legislation in the United States can be understood as a result of belief changes about the implications of bans on the alcohol market. While Prohibition received support from a fraction of the population that held moral views against alcohol consumption, their beliefs that such bans could be implemented effectively and would reduce rather than increase crime contributed to their zeal. These beliefs changed rapidly, however, as communities experienced sharp increases in crime following the implementation of Prohibition. Many former supporters of the policy then found themselves in a situation similar to that of John D. Rockefeller, himself a radical prohibitionist, who recognized such a tension in the late 1920s:

When Prohibition was introduced, I hoped that it would be widely supported by public opinion and the day would soon come when the evil effects of alcohol would be recognized. I have slowly

---

1"From a formal point of view, one cannot distinguish between an individual’s dislike for having his grounds ruined by factory smoke and his extreme distaste for the existence of heathenism in Central Africa... I merely want to emphasize here that we must look at the entire system of values, including values about values, in seeking for a truly general theory of social welfare." Arrow (1963, p. 18) In Arrow’s terms, an individual who performs a private activity which another individual considers immoral will, as a result, impose an externality onto him, out of the latter’s regard of the former’s action as immoral. Thus, for example, there is widespread agreement across individuals regarding the immorality of murder, but widespread disagreement regarding the morality of abortion.

2The salience of moral issues in the political agenda could be a result of convexity of preferences over them, as in Kamada and Kojima (2010), or because they are strategically exploited by an interest group, as in Baron (1994). In the context of Prohibition in the U.S., the latter seems a better description of the process leading to the adoption of Prohibition.
and reluctantly come to believe that this has not been the result. Instead, drinking has generally increased; the speakeasy has replaced the saloon; a vast army of lawbreakers has appeared; many of our best citizens have openly ignored Prohibition; respect for the law has been greatly lessened; and crime has increased to a level never seen before. (John D. Rockefeller, quoted in Okrent (2003, p. 246-247))

In this paper I study the relationship between policy reform and social change, and argue that ex-ante uncertainty about the effects of radical changes in society’s legal standards, coupled with the ability of individuals to learn about the effects of those policies, can be at the heart of the dynamics of social change, through a feedback between the effects of policies and changing attitudes in response to their effects, modulated by the endogenous extent of enforcement of those same policies. More specifically, I exploit the Prohibition experience of the 1910s-1930s to investigate the extent to which support for different types of bans is determined by the interplay between moral views and beliefs, and how this support changes as beliefs evolve as a result of learning from the outcomes of those policies.

In fact, as a methodological contribution, I argue that the mechanism proposed in this paper may have relevance outside the experience of Prohibition to understand the evolving attitudes towards moral issues, and more generally to think about the forces shaping social change. Attitudes towards Catholics in the 19th Century U.S., towards the role of women around the mid 20th Century, towards blacks in the South after the Civil War and after the Civil Rights Movement, or more recently towards Muslims in Western countries, for example, could be better understood by studying how the enforcement of policies targeted towards specific groups has effects that change collective preferences over those policies, endogenously feeding back into changes in policy choices, and in individual attitudes in the long run.

With this purpose, I develop and estimate a dynamic structural model of Prohibition enforcement and crime, where heterogeneity in moral views and beliefs interplay, and have observable and unobservable components. Learning is rational, and communities decide the enforcement margin of Prohibition through a collective decision. Law enforcement shifts the distribution of crime, and individuals update their beliefs about the effects of Prohibition by observing homicide rate realizations. Because law enforcement is endogenous to preferences and beliefs, the speed of learning by rational agents is affected not only by their priors, but also, indirectly, by the distribution of moral views giving rise to such collective choices of law enforcement.

I estimate this model by Conditional Maximum Likelihood, using a dataset of U.S. cities during the period 1911-1936, when the country experienced a Prohibitionist wave which reached Constitutional status, and focus on the homicide rate, the drunkenness arrest rate, and police expenditure as the main observable outcomes. I start by showing that crime and law enforcement during Prohibition presented a rise and fall pattern, and that the alcohol market contracted and rebounded quickly thereafter (see figures 1-2). Then I document how these patterns differed between cities with varying moral preferences, by using observable variation in the distribution of religious ascriptions and other demographics: drier (i.e., more favorable to Prohibition) cities experienced initially higher levels of law enforcement, while wet (i.e., less favorable to Prohibition) communities observed higher increases in criminality and larger changes in public support for the policy.
The estimated structural model explains a large fraction of the variation, both across cities, and over time, in the choices of policing expenditure in cities, the observed evolution of criminality measured through the homicide rate, and the alcohol-market dynamics. With the model I also estimate the extent to which Prohibition as a legal standard, and Prohibition enforcement, were responsible for the increase in criminality observed during the period. Prohibition was associated with an average homicide rate increase of 15 to 20%, while it was unable to shrink the alcohol market. At its lowest point, around three years into Prohibition, the effective alcohol supply fell by around 35%, but rebounded quickly thereafter. Moreover, I estimate that the Prohibition-related homicide rate was increasing in the level of law enforcement. Relatedly, cities in smuggling areas had a lower potential for crime to develop under Prohibition, and I argue this was due to the reduced constraints faced by the black alcohol market in those areas.

The structural model also allows for the estimation of several moments of the joint distribution of moral views about Prohibition, and prior beliefs about its effects. I find that beliefs were extremely optimistic across the distribution of moral views, so that the variation in moral views across cities was larger than the variation in initial beliefs. Although people had strong opinions about alcohol Prohibition at its outset, there was not much disagreement about its effects. Nevertheless, the estimated correlation between moral views and beliefs is large, implying that drier individuals held even more optimistic prior beliefs about the effects of the policy.

I conclude with a series of counterfactual exercises based on the structural model, which illuminate the key interactions taking place during Prohibition. I find that local policy was highly responsive to community preference changes. As a result, a more polarized society would have learned faster, but also would have observed higher crime increases during Prohibition. Communities would have responded to Prohibition by offsetting it with reduced law enforcement choices if prior beliefs had been less optimistic; this would have reduced the crime spike of the 1920s, but would have limited the speed of change in public opinion. Finally, in an exercise where local decision-making power is shifted away from the median voter, the increased misalignment between the community's distribution of preferences and the equilibrium law enforcement choice alters the speed of learning by changing the informativeness of the crime signals.

1.1 Related Literature

This paper is related to several research areas. The first studies the determinants of civil liberties. To my knowledge, Lagunoff (2001) is the only work which directly addresses the question of why democracies are able to sustain civil liberties for minorities. According to his argument, when a majority is likely to become a minority in the future, it will have incentives in the present to weaken the enforcement technologies available, which could otherwise be used against them in the future. Political Scientists, on the other hand, have stressed that the salience of moral issues is relevant to understand the extent of civil liberties, because it determines the degree to which the legal standards adopted will respond to interest group politics (Haider-Markel and Meier (1996)). Indeed, through the political system, different practices are prohibited or restricted based on moral motivations alone. In autocratic societies, rulers and elites directly impose their moral views upon the community; in democracies, majorities can impose restrictive legal standards upon minorities.
through the ballot box. Legal restrictions on individual liberties are of economic and political importance for several reasons. First, they directly have welfare implications over both individuals who favor and disfavor the prohibition. More interestingly, they often have potentially uncertain side effects. The imposition of a restrictive legal standard creates dissatisfaction in a subset of the population, leading to non-conformist behavior, political mobilization, unrest, or violence.

As a result, *de jure* prohibitions require concomitant *de facto* enforcement. Because they are prone to widespread loopholes, enforcing restrictions on the behavior of individuals requires costly monitoring and willingness to enforce, both by the community and by its enforcement agents. In fact, within the literature on “crime and punishment” pioneered by Becker (1968), concerned with the understanding of the determinants of crime enforcement and the effects of law enforcement on the equilibrium levels of illegal activity, this paper highlights that what society defines as crime is endogenous, and that, as a result, punishment is a social choice. These considerations have been overlooked in the literature, and suggest that agreement about punishment within society, and social learning about its costs and benefits, might be important to understand the success of alternative policies.

In this paper, the main channel driving public opinion and law enforcement outcomes is the interaction between beliefs and moral views, making it close to the research on policy and rational learning. Landier et al. (2008) and Alesina and Fuchs-Schundeln (2007) study how ideological differences have affected beliefs about capitalism. Sargent et al. (2006) develop a statistical model about monetary policy in the U.S., where policymakers endogenously learn about the Phillips curve. Buera et al. (2010) is a recent example of structural estimation of a learning model, where policymakers update their beliefs about the merits of market oriented versus interventionist policies by observing their neighboring countries’ outcomes. A theoretical paper in the same spirit is Mukand and Rodrik (2005), who argue that experimentation and imitation might explain why, over the last decades, countries have converged in the adoption of policies, but not in economic performance. Strulovici (2010) is also an important recent contribution, which studies the incentives for policy experimentation in a dynamic voting framework, in which incentives for experimentation are limited by the trade-off between learning about the effects of policies and the pivotality of voters.

Finally, this paper also contributes to the literature on crime (See Dills et al. (2008) for a recent survey where the authors conclude that the most robust correlate of crime is the prohibition of drugs). The literature has stressed factors like the age composition of the population (and abortion), the deterrence effects of incarceration, access to firearms, investment in policing, inequality, or the economic cycle (Levitt (2004); Donohue and Levitt (2001); Dills et al. (2008)). Miron (1999a) and Goldstein (1985) stress the main channel I explore in this paper, where non-conformism and law enforcement over activities involving traded commodities create the potential for violence and corruption to arise as salient side effects. Competition for black market rents, unavailability of institutionalized channels of dispute resolution, and the use of coercion by law enforcers, all create incentives for crime and corruption to develop (Miron (1999b)). Thus, tightening law enforcement can magnify the effects of the prohibition on crime. It drives out the marginal producers (which are less likely to engage in criminal behavior), weakens social norms sustaining peaceful dispute resolution among criminals, and crowds out resources for overall crime enforcement (Becker (1968); Miron (1999b)). The literature has mostly focused on reduced-form or instrumental variables strategies, whereas I explicitly model the endogenous relationship between law enforcement choices and crime.
that arises in the context of Prohibition, highlighting the role of rational learning and beliefs.

The rest of this paper is organized as follows. Section 2 presents a historical overview of the Prohibition experience in the United States during the early decades of the Twentieth century, and discusses its institutional and political background. Section 3 then presents and discusses the data collected and used in the paper. Based on the historical discussion, section 4 subsequently presents reduced-form results, which guide the development of the model presented in section 5. Section 6 proceeds with the estimation results from the structural model, and presents some counterfactual exercises. Finally, section 7 concludes.

2 Prohibition: A Historical Overview

2.1 Prohibition Politics

Nation-wide alcohol Prohibition in the United States was written into the Constitution as the 18th Amendment in January 1919, and repealed from it just fourteen years later, as the 21st Amendment, in December 1933. Given the constitutional supermajority requirements to amend the U.S. Constitution, such a policy reversal is striking. The increase in criminality during the Prohibition period, best illustrated by figure 1, was as striking, and a first-order reason why public opinion had such a radical swing in such a short period of time. Alcohol Prohibition, though, was not a sudden appearance; it was the endpoint of a prohibitionist wave with origins dating as far back as the 1870s, when a group of Ohio women organized the so-called Temperance Crusade, which would later give rise to the Women’s Christian Temperance Union (WCTU).

Prohibition was introduced staggeredly across counties and states through a gradualist political strategy of religiously motivated temperance groups, closely related to the Baptist, Methodist and Evangelical churches, and composed mostly of native-born whites and women (Sinclair (n.d.); Okrent (2010)). The two most prominent were the WCTU, and the Anti-Saloon League (ASL). Both developed a nationwide organizational structure, but the ASL took the lead in the beginning of the Twentieth century. Initially these groups were not a majority of the population. Their political success was due to their pivotal character in the competitive context of bipartisan politics, based on strong campaigning and lobbying in state legislatures, towns, and cities, and on the intensive use of referenda initiatives. Republicans and Democrats were frequently so evenly divided that a switch of the temperance vote could easily decide local elections. Prohibitionist groups were able to become pivotal even in the within party races of the Democratic-dominated South. Their persistence in lobbying also was important because Prohibition was not an issue that politicians paid much attention to at the time

Key to the political success of the drys was their strategic avoidance of aligning with either party. While the ASL was relatively antagonistic to Northern Democrats whose constituencies were mostly

---

3 Constitutional amendments require approval by two thirds of the vote in both the House and the Senate, and a plurality of the vote in either both chambers of at least three fourths of the State Legislatures, or in at least three fourths of State Constitutional Conventions.

4 Talking about the 18th Amendment, Sinclair (n.d., p. 182) argues that “... boredom played some part in the passage of the amendment. The members of Congress were sick of being badgered by the Anti-Saloon League and their dry constituents.”
in large urban areas, it was much closer to Southern Democrats for whom Prohibition was another channel for social and political control of blacks (Asbury (1950, p. 93), Sinclair (n.d., p. 182)). There was disagreement on the issue within Democrats in the South too, as a faction of the party believed that allowing the Federal government to make decisions regarding Prohibition could be the first step to further undermine Southern autonomy (Szymansky (2003)). An indicator of the lack of partisan alignment on Prohibition is the House roll call on the 18th Amendment; 64 Democrats and 62 Republicans voted against, while 140 Democrats and 138 Republicans did so for Prohibition. A second important element to explain the success of the dry campaign was its gradual approach. Local option measures were followed by state-wide legislation, so that right before the 18th Amendment was adopted, almost 80% of U.S. counties were already under some form of Prohibition, starting from no more than 15% in 1900. Figure 3 shows the dates of state-adopted of Prohibition legislation.

The rise of prohibitionist attitudes in the U.S. was part of the so called “Progressive Era”, a much broader set of social and political changes taking place in the late Nineteenth and early Twentieth Centuries, associated with the rapid expansion of State capacity throughout the country. The establishment of the income tax under the 16th Amendment and the enfranchisement of women under the 19th Amendment also were part of the expansion of the role of the State in society. In this perspective, Prohibition expanded the role of the State into the private activities of individuals. This required an unprecedented involvement of the churches in politics, fueled by a context of rapid social change and urbanization, which was increasing the heterogeneity of the American society. On the dry side, priests moved from claiming the sinfulness of drinking, to advocating explicitly prohibition legislation (Isaac (1965, p. 263)). On the wet side, it was about “whether or not the American people were going to hand over to government the paternalistic power to regulate lives and habits” (Kyvig (1979, p. 51)).

2.2 Law Enforcement

Before Constitutional Prohibition, enforcement of the alcohol laws in states under Prohibition was usually mild. In dry communities it was redundant, while in wet communities it was relatively ignored. A large share of alcohol consumption took place in saloons and other public spaces, which made public intoxication a widespread phenomenon (See, for example, Blocker (2006); Stayton (1923)). Prohibitionist associations were concerned about the social consequences of saloons, and arrests for drunkenness were seen as a key indicator of successful enforcement of dry laws. But loopholes were abundant and often overlooked (Franklin (1971)). The biggest loophole was probably interstate shipping of alcohol into states under Prohibition. As a response, the ASL lobbied intensively until it achieved the passage of the Webb-Kenyon Act in 1913, banning interstate shipping of alcohol into dry states. Although later it was practically unenforced, at the time of its passage this law was very controversial. President Taft vetoed it, and Congress subsequently overrode his veto. Wets, backed by the Brewers Association, argued the law violated the First Article of the U.S.

---

5 Okrent (2010, p. 33) quotes a WCTU strategist who, being asked why alcohol was inconvenient, gave the following account: "...selling in prohibited hours, gambling, selling to intoxicated men, rear rooms, unclean places, invading residential districts, the country saloon, the social evil, selling to minors, keeping open at night, brewers financing ignorant foreigners who are not citizens, the American bar, brewery-controlled saloons, cabarets, Sunday selling, treating, free lunch, sales to speakeasies, bucket trade, signs, screens, character of the men, too many saloons".
Constitution, but the Supreme Court later upheld it.

At the same time, although the passage of the 18th Amendment and its enforcement law (the Volstead Act)\(^6\) appeared as highly restrictive by banning any liquor with more than 0.5% alcoholic content, Congress did not make large appropriations for its federal enforcement. In fact, the Amendment established concomitant enforcement by the local, state and federal levels, so Congress, expecting cooperation from local and state policing agencies and general compliance with the law, created a modest federal enforcement organization (Kyvig (1979, p. 23)). The weakness of federal enforcement is best exemplified by the constant changes in Prohibition administration during the 1920s\(^7\).

Table 1 presents the main Federal Prohibition law enforcement outcomes during the 1920s. Trends are very similar across the four main U.S. regions, and suggest that enforcement intensity peaked around 1928. Early during national Prohibition, given the initial absence of domestic producers, most of the supply of illegal liquor came from international smuggling (Okrent (2010)). Over time, local production based on illegal distilleries and stills caught-up with demand. Nevertheless, the number of distillers and fermenters seized fell sharply in the later Prohibition years, which suggests a sharp fall in the enforcement activities against producers. The number of killed or injured agents during enforcement activities shows that the 1927-1930 period was particularly violent, and that subsequently few risky law enforcement activities took place. The 1927-1930 period coincides with the years in which Prohibition administration was under the Bureau of Prohibition, in what historians have acknowledged as the last attempt form the federal government to control the liquor trade, in response to the fall in state and local law enforcement throughout the country. Indeed, by 1928 several states had already repealed their own enforcement legislation. To have an idea of the limited extent of law enforcement at the federal level, notice that in 1929-1930, total liquor seizures in the U.S., including spirits, malts, wines, cider, mash, and pomace, were approximately 74 million gallons. Compared to the 3,375 million gallons of booze which, according to Okrent (2010, p. 202), were produced and distributed annually by Max Hoff, an illegal producer in Pennsylvania, the federal enforcement looks almost irrelevant.

In fact, most of the law enforcement, in practice, relied on local efforts. This was not only because of the inherent difficulties in enforcing alcohol restrictions throughout the country, which limited the federal law enforcement strategies to infrequent raids and a focus on some particularly troublesome areas, but also because of the inefficiency of the federal agency. Complaining about this issue in 1926, Colvin (1926, p. 497) argued that, “Although the United States had adopted a national standard throughout the nation, the administration of the law so perverted this objective as to make enforcement substantially a matter of local opinion because it was administered to so large

---

\(^6\)President Wilson also vetoed the Volstead Act, and his veto was also overridden by Congress.

\(^7\)Originally, the Volstead Act created the Prohibition Unit as a department of the Bureau of Internal Revenue, with Prohibition Directors in each state. The Coolidge administration avoided dealing with the Prohibition problem throughout, and in 1925, there was a sharp reduction in the size of the Prohibition Unit (Colvin (1926, p. 495)). The critical situation regarding corruption and venality within it resulted in a reform of Federal Prohibition administration under the Prohibition Reorganization Act of 1927. This act created the Bureau of Prohibition, ascribed to the Treasury Department, putting its employees under the Civil Service and creating 27 Prohibition Districts (Schmeckebier (1929), Schmekebier (1923)). Finally, in 1930 the Prohibition Bureau was transferred to the Justice Department, but at this point, “...as useful as these congressional steps may have been... the enforcement effort had acquired a dismal reputation and doubts as to whether Prohibition could possibly be effective had become deeply engrained” (Kyvig (1979, p. 32)).
a degree by men owing their appointment to local political influences and subject to local political pressures... it was the worst form of local option - the option of the local politicians to determine the extent to which the law should be enforced-, politicians, many of whom were personally wet, others of whom wanted to placate a wet element in their constituencies, and all of whom belonged to political parties which sought wet votes as well as dry ones”. While a dry such as Colvin saw the problem in the ineptitude and corruption of enforcers, a wet such as Tydings would argue that “If moral force... does not make them stop, physical force must be used. Where is the physical force to come from? Plainly, in a nation of 120 million people, scattered over an area of 3 million square miles, the force must be predominantly supplied by the local enforcement authorities... but the police, the courts and the juries are the servants and reflectors of local sentiment”(Tydings (1930, p. 125)).

Thus, the degree of law enforcement of Prohibition was responsive to the local demand for both Prohibition and alcohol, and elected authorities were agents of both groups. This seems to have been true not only during Constitutional Prohibition, but also during state-level Prohibition. Franklin (1971), for example, quotes a local judge in dry Oklahoma claiming that a candidate for sheriff would not possibly be elected, if it were known that he intended to enforce Prohibition. In the same way, judges and juries tended to be lenient in their decisions regarding Prohibition violation cases (Szumansky (2003, p. 184), Kyvig (1979, p. 25), Tydings (1930, p. 127)). Judicial leniency was even institutionalized through the so-called “bargain days”,8 which arose in response to the courts’ congestion created by the overwhelming number of violations of the Volstead Act. In fact, initiated criminal prosecutions in federal courts for violations of Prohibition increased from slightly more that 100 per million inhabitants in 1920, to almost 500 in 1925, which made up 80% of all criminal prosecutions9.

If law enforcement varied as a function of local preferences, the effects of Prohibition also varied between communities. This is acknowledged by a Commissioner traveling around the State of New York in 1930 who argued that the problems varied between and within states, particularly between the rural and urban areas10. According to Kyvig (1979), Scandinavians in Minnesota continued to drink, while Idaho, Oregon, and Washington had come to accept Prohibition. Los Angeles and even San Francisco had large dry constituencies, and relatively dry areas ran from California to Texas. Louisiana, on the other hand, was extremely wet and law enforcement relied almost exclusively on federal authorities. In the rest of the South, Prohibition was enforced particularly on blacks. Finally, in the large wet cities of the Northeast such as Detroit, Cleveland, Pittsburgh, Boston and New York, Prohibition was largely unobserved, and weakly enforced, particularly after the second half of the 1920s.

The weakening of law enforcement took place not only by a reduction in policing and prosecution, but also through the repeal of state enforcement legislation. The most prominent case was that

---

8Violators would plead guilty and be charged a small fine.
9I collected the data on judicial prosecutions at the judicial district level for the period 1915-1933 directly from the Attorney General Annual Reports.
10New York City presents a problem quite distinct from the up-state section, and the border region presents an entirely different situation... the problem varies as the population is homogeneous or heterogeneous... throughout the rural and smaller cities... there is a greater respect for the law and established order” (Wickersham-Commission (1928-1931b, Box 13-2, Prohibition Survey of New York, p.2)).
of New York, which very early on, in 1923, repealed the state enforcement law. Alfred Smith, the Democratic Presidential candidate in the 1928 election, was then the Governor of New York. The repeal by no means a consensual decision, and in fact, many dry organizations lobbied Smith to veto it. It was a difficult decision because, although openly wet, alienating the dry vote could prove costly for his future political career. In his own words, “Some seem to think that my approval [of the repeal] will mean the preservation of American Institutions. Many others impeled by equally patriotic motives seem to feel that my approval will be destructive of American government. Obviously, both cannot be right...” (Smith (1923, p. 601)).

2.3 Repeal

The early repeal of state enforcement legislation in New York was driven more by the morally anti-Prohibitionist character of its large share of urban population than by a rise in criminality, which by that time, had not yet peaked. The shift in public opinion in other regions of the country took place at a slower pace, and more in response to the observable increase in criminality. Initially dry individuals, who were morally compelled by Prohibitionist reasoning, could not avoid acknowledging the adverse consequences that the policy was having.

The rise in crime and undermining of the rule of law was not homogeneous across the country, and as a result, neither was the fall in support for the policy. The Democratic party, which had been out of power throughout the 1920s, managed to capture most of the rise in anti-Prohibitionist sentiment. In the 1928 Presidential election this hurt Al Smith, but in the 1932 campaign it played in favor of Franklin D. Roosevelt. The distribution of public opinion did shift massively against Constitutional Prohibition, and opposition became better organized. The Association Against the Prohibition Amendment, for example, began its advertising campaigns in 1928, focusing on providing information about the ill-effects of Prohibition. In 1929, the Women’s Organization for National Prohibition Reform was founded with the same intentions. Nevertheless, even after the repeal of the 18th Amendment, six states remained dry. Among the rest of the states, some instituted systems of “state operation”, in which the state directly controlled the distribution of alcohol, while others just imposed some regulation over a free market (Harrison and Laine (1936, p. 43)).

3 Data and Summary Statistics

3.0.1 Data on Criminality

Criminality was the main source of concern and learning about Prohibition for the public. The homicide rate is the variable for which most comprehensive information is available, and one for which measurement error is likely to be very limited. Thus, I collected information from the Mortality Statistics published yearly by the Bureau of the Census, reporting the number of non-traffic-related homicides for a sample of U.S. cities. I complemented this information with the homicide data

---

11These were Alabama, Kansas, Mississippi, North Carolina, North Dakota and Oklahoma. Nonetheless, all of these, except Alabama and Kansas, allowed for the sale of beer (Kyvig (1979, p. 188))
reported in the Wickersham Commission documents, finally putting together yearly data for the period 1911-1936 and a sample of up to 93 cities. Data on drunkenness arrests, on the other hand, is very detailed and covers a total of 573 cities for the period 1910-1929\textsuperscript{12}. Finally, the Federal Bureau of Investigation began compiling and publishing its Uniform Crime Report (UCR) in 1930, which contains yearly city-level data on murders and other offences reported to the authorities. Offences include robbery, assault, burglary, larceny and auto theft.

3.0.2 Law Enforcement Data

Law enforcement is a difficult concept to measure because it depends on the discretion of the enforcer, and thus, is necessarily unobservable. Moreover, measuring law enforcement through its outcomes is problematic; an increase in liquor stills seized, for example, could be explained by an increase in Prohibition enforcement on a constant level of illegal alcohol production, or by a reduced level of law enforcement which allows for illegal production to increase. Because a great deal of Prohibition enforcement, and all of local crime enforcement, was decided and implemented at the city level, I focused on collecting data on city public finances, and specifically, on police expenditure. I use the Financial Statistics of Cities published yearly by the Bureau of the Census, which report disaggregated data on city public finances for cities with populations above 30,000 (around 250 cities), and obtain data on total city public expenditure and investment, police expenditure and investment, and all protection expenditure and investment (all protection includes police, fire and other expenditure), for the period 1911-1936. I computed 1913-constant prices expenditure data by using the U.S.-wide CPI as of June of each year as the deflator\textsuperscript{13}.

3.0.3 Demographic and Religious Data

City and county-level data on demographic characteristics are taken from the decennial population censuses. I focus on the age distribution, the ethnicity distribution\textsuperscript{14}, and total population, from the 1910-1940 Censuses. Given the strong relationship between ethnicity and religiosity with attitudes towards the liquor problem, I use religious ascription data from the decennial Censuses of Religions (1906, 1916, 1926, and 1936), to capture heterogeneity in moral views about Prohibition. I aggregated religious ascriptions in the following nine groups, directly from their names: Baptist, Eastern Orthodox, Evangelical, Jewish, Mormon, Lutheran, Methodist/Episcopal, Catholic, Presbyterian, and other. The consensus amongst historians is that Baptist, Evangelical, Mormon, Methodist, Episcopal and Presbyterian communities held the strongest views in favor of Prohibition, while Catholic, Orthodox, Jewish and Lutherans had much more favorable positions regarding alcohol consumption (See Foster (2002); Lewis (NA); Szymansky (2003)). I refer to the former as “dry”, and

\textsuperscript{12} The data on drunkenness arrests contained in the Wickersham Commission papers appears to have been originally compiled by the World League. Dills et al. (2005) use this source, covering a shorter time period, together with an alternative source compiled independently by the Moderation League. Both series appear to be highly correlated, so I restrict attention to the World League data, which covers the whole 1911-1929 period.

\textsuperscript{13} Data for the years 1914 and 1920 is unavailable. For the balanced panel estimations below, I use the interpolated values (1913-1915 average for 1914, and 1919-1921 average for 1920) for these two years.

\textsuperscript{14} I focus on the distribution of the population between native white, foreign white, and black individuals.
to the latter as “wet” religions. I then computed the share in each religion directly as the number of adherents divided by the total number of adherents to any religion in the city (or county).

3.0.4 Public Opinion Data

To measure public opinion about Prohibition, I collected electoral returns data on referenda on alcohol-related issues for the different states, taking place during the 1900s-1930s. These referenda were usually ballot measures proposed to the citizens to approve or repeal liquor laws, or amend the state constitutions. In states where local option was in place, county or city-level referenda had the purpose of allowing or forbidding the sale of alcohol. When submitting the 21st Amendment to the states, the U.S. Congress determined that Constitutional Conventions should be elected in the different states to decide over the issue, and candidates should run in either a dry or a wet slate (Brown (1935)). All of the referenda returns allow me to directly compute the fraction of (anti-Prohibitionist) wet vote, which I use as a proxy of wet support. Almost all of the electoral returns data is available at the county level, except for referenda in the states of Connecticut and Massachusetts, for which city-level data was reported. Overall, I have referenda election returns for 2,083 counties.

3.0.5 Legislation Data

Alcohol-related legislation across states comes from three main sources: the Anti-Saloon League’s 1916 Yearbook and the information in Szymansky (2003), and in Cherrington (1920). The latter source was in particular very useful since the author makes a state-by-state compilation of all of the dry legislation up to 1920, detailing the time of its passage and/or repeal, and providing a brief description of it. Based on these sources, I coded a state-level variable for the number of dry laws in place in each year, an indicator variable for being under Prohibition (either state-level or federal-level), and an indicator variable for having a Prohibition enforcement law in place.

3.1 Summary Statistics

Table 2 reports population-weighted summary statistics for the main variables used in the paper, summarizing the available information for up to 340 cities (counties for the referenda election returns data), and disaggregating the sample in the four main U.S. geographic regions. The table presents the baseline distribution of religious ascriptions and demographics, together with data on legislation. It also includes summary statistics for the different outcomes of interest, comparing average values in the 1910s and 1920s.

For the religious distribution, I present summary statistics from the 1916 Census of Religions. As expected, Southern cities were heavily Baptist and Methodist relative to the rest of the country (29% and 24% respectively). The South was also less Lutheran and Catholic. Indeed, Catholicism was

---

15 The main caveat here is that differences in turnout rates might differ systematically between Prohibitionist and anti-Prohibitionist voters, not reflecting the true distribution of political preferences in the community. For an empirical model of turnout on alcohol-related referenda, see Coate and Conlin (2004).
concentrated in the Northeast and Midwest, where more than half the adherents in the sample belong to this religion. Evangelicals were mostly concentrated in the Midwest, while Mormon communities were mostly found in the West. In fact, with almost a 50-50 split between dry and wet religions, the Western cities present the more uniform distribution of religious membership. In contrast, religious membership in Southern cities was heavily skewed towards dryness, while in the Midwest and Northeast wet religions were majoritarian.

Looking at the basic ethnic composition across regions from the 1910 Population Census, 26% of the population in the Southern cities in the sample was black, in sharp contrast with all other regions where the black population was between 1.3 and 3.1 percent. The foreign white population was especially prevalent in the Northeast, where 32% were whites born outside the United States, as compared to only 7% in the South. In the Midwest, on the other hand, almost three quarters of the population was native white.

A look at the outcome variables reveals that real per capita expenditure in police was significantly larger in the 1920s than in the 1910s, with an average increase of around 0.3 dollars. Northeastern cities had the highest levels of expenditure in both decades, but Southern cities experienced the largest average increase. Although per capita expenditure in police rose, the data on police expenditure as a share of total city expenditure reveals a fall everywhere, due to the fast increase in public spending in other categories during these Progressive Era decades. Cities in the West had the lowest police shares (around 8%). While per capita policing was lowest in the South, Southern cities had the highest share of their budget allocated to police (11 – 12%). The average behavior of the data on drunkenness arrests reveals considerable differences between regions. In Southern cities, average arrests were very similar in the 1910s and 1920s. In contrast, cities in the West do show a large fall in arrests for drunkenness between both decades, falling from 22.5 to 13.9 per 1,000 inhabitants. Although arrests in the Midwest and Northeast also are somewhat lower in the 1920s, the fall is not as large.

The homicide rate, on the other hand, shows significantly higher levels in the 1920s in all regions, and large level differences across them. While homicide rates were on average 5.3 per 100,000 in Northeastern cities during the 1910s, they were almost five times higher in the South during the same decade. The variance of the homicide rate was also much larger in the South. It is also worth noticing that the smallest average increases in the homicide rate took place in the West, where it only increased from 9.8 to 11.6.

Support for Prohibition, as measured by the electoral returns on alcohol referenda, was higher in the South and the West, where the wet vote shares were 0.46 and 0.45 on average, while it was slightly above 50% in the Midwest and the Northeast. A comparison of these numbers between decades reveals the striking shift in public opinion; wet support was around 20 percentage points higher in the West and Midwest, 30 percentage points higher in the Northeast, and 10 percentage points higher in the South after Prohibition. Interestingly, the South showed the smallest increase in wet support, while, despite its higher initial anti-Prohibitionism, Northeastern cities experienced the largest average shift against Prohibition.
4 Some Reduced Form Results

I begin the empirical analysis by focusing on three first-order sources of variation in the effects of Prohibition. Differences in the timing of its adoption across states, in preferences over the legal standard (i.e. moral views about alcohol consumption), and in state-level legislation and its enforcement both at the local and federal levels. I focus on three outcome variables: as a direct measure of criminality, the homicide rate shows a large increase, happening with some delay after the introduction of Prohibition, and reaching its highest levels around the mid 1920s. Crime increases were larger in cities with bigger potential alcohol markets and populations less inclined to the policy. However, they were similar between cities facing different initial crime levels. By looking at the drunkenness arrest rate, I document a drastic contraction of the alcohol market right after the introduction of Prohibition, but a steady and relatively fast recovery. Neighboring markets reduced the extent of contraction in alcohol consumption, and the time path was remarkably similar across different cities. Finally, there is evidence of a steady increase in law enforcement following the introduction of Prohibition, with a subsequent fall starting in the late 1920s. The early increases in law enforcement were faster in cities with constituencies more favorable to Prohibition, but for late Prohibition years, these cities show lower spending in policing. Subsequently, I look at changes in public opinion regarding Prohibition by exploiting electoral data on liquor referenda, and document a non-monotonic relationship between changes in public opinion and overall moral views of cities' constituencies: communities with intermediate levels of initial support towards the policy saw the largest shifts in public opinion against Prohibition.

4.1 Crime, Law Enforcement, and the Timing of Prohibition

A natural first approach is to compare outcomes before, during, and after repeal of Constitutional Prohibition. Figure 1 shows that the advent of Prohibition saw a sharp increase in crime, here measured by the homicide rate (although a mild, positive pre-trend can be observed since the early 1910s). Nonetheless, it also suggests that the difference was not constant throughout the fourteen years after its adoption; the homicide rate increased rapidly during the early years of Constitutional Prohibition, and slowly started to fall back to pre-Prohibition levels around 1926.

Observed arrests for drunkenness are the equilibrium outcome of alcohol demand, alcohol supply, and intensity of arrest enforcement. Their evolution captures changes in all of these components. Figure 2 presents the population-weighted average per-capita drunkenness arrest rate for the 255 U.S. cities for which this variable is available throughout the whole 1911-1929 period. Its sharp fall started well before Constitutional Prohibition was adopted. It fell to around 39% of its initial level (from around 18 to only 7 arrests per 1,000) in just a few years. On the other hand, it was precisely in 1920, the year when the 18th Amendment entered into force, that drunkenness arrests started bouncing back at an even faster rate. They finally converged to around 83% of their average initial level, at a time when federal Prohibition was still in place. The breaks in both the homicide rate and the drunkenness arrest rate series do not appear to match the introduction of Constitutional Prohibition. This suggests differential short-run and long-run effects of Prohibition, and the relevance of state-level Prohibition, which, as mentioned in section 2, occurred staggeredly
across states during the first two decades of the century.

In fact, throughout the 1910s arrests have a sharp fall in every city, but at different points in time across cities in different states. The fall appears to be highly correlated with the timing of adoption of state-level Prohibition. Figure 3 presents the dates of adoption of state-level Prohibition. It shows how the Prohibitionist wave moved across the United States during the 1910s, up until the introduction of nationwide Prohibition with the 18th Amendment. 

In the context of alcohol Prohibition, time under the policy is a convenient reduced-form way to look at its time-varying effects for several reasons. First, because of the alcohol supply dynamics; after Prohibition is adopted, the legal market for alcohol is closed on impact. This implies a large negative shock on the availability of liquor. The black market requires time to develop smuggling networks and establish hidden production facilities. Moreover, because crime is a necessary input into the production and trade of any illegal commodity, costly and time-consuming investments are also necessary for the development of criminal organizations supporting the illegal market. Finally, law enforcement was a key channel through which Prohibition had an impact on the development of criminality, and equilibrium law enforcement depended on the community’s beliefs about the policy. The evolution of these beliefs over time was also a dynamic force shaping the time-varying effects of Prohibition as a legal standard. Thus, Prohibition is likely to have varying short-run and long-run effects. To obtain an estimate of the overall effects of Prohibition, a comparison of cities which have experienced similar lengths of time under the policy is needed.

To take a first look at short-run and long-run effects of Prohibition, I start by estimating fixed-effects models of the form

\[
y_{ct} = \alpha_c + \beta_t + \sum_{\tau=1}^{k} \delta_{\tau} D_{ct} + \gamma' X_{ct} + \epsilon_{ct}
\]

where \(c\) indexes cities and \(t\) indexes years. \(y_{ct}\) can be either the homicide rate, the drunkenness arrest rate, or police expenditure, for which I look at two alternative measures: Police expenditure as a share of total city public expenditure, and per capita police expenditure. The \(\alpha_c\) are city-specific effects, the \(\beta_t\) are year-effects, and the \(D_{ct}\) are indicator variables for each cumulative number of years under Prohibition. The vector \(X_{ct}\) includes a constant, the log of population to capture any scale effects, and time-varying effects for border and state-capital indicators. The focus of Equation (1) is in the estimates of \(\delta_{\tau}\), the time-varying effects of Prohibition. Since this model looks only at within-city variation over time, the \(\delta_{\tau}\) can be interpreted as the average-across-cities difference in \(y_{ct}\) relative to the city average, when a city has been under Prohibition for \(\tau\) years. Standard errors reported are robust to arbitrary heteroskedasticity and clustered at the city level to adjust for arbitrary within-city correlation over time. Because of the strong trend in the police expenditure

---

16 In figure 3, Kansas, Maine and North Dakota are not shown because these three states were already under Prohibition since the late 19th century. Kansas adopted Prohibition in 1880, Maine in 1884, and North Dakota in 1889 (at the same time it acquired statehood). Kansas and Maine had already been under Statewide Prohibition in the mid-1800s during the first Prohibitionist wave.

17 In the sample \(\tau\) runs up to 55, given that Kansas was under Prohibition since 1880. Because only very few cities experienced Prohibition for more than eighteen years, I restrict \(k\) to be 19, and leave observations with more than nineteen years under Prohibition as part of the omitted category.
data, I also ran "random trend" models for some specifications for this outcome variable, allowing for city-specific linear trends.

I present regressions for two alternative samples, labeled as B and C\(^1\). B is a balanced sample of the 66 cities for which complete data is available for the whole period 1911-1936, which will be the sample used for the structural estimation in sections 5 and 6. Sample C is an unbalanced panel excluding cities for which there are less than ten years of data for drunkenness arrests or police expenditure, or less than eight years of homicide rate data. Thus, \( B \subset C \subset A \). The complete regression results can be found in Appendix 4. For brevity and ease of illustration, the left panel in figure 4 graphs the estimated \( \delta_r \)'s of the baseline specification with no year effects. It nicely shows how the homicide rate is relatively unresponsive for the first few years after a city has been under Prohibition, and then trends upwards until around the 10th year under Prohibition. The homicide rate then starts slowly falling back to a level similar to the pre-Prohibition average. The set of cities experiencing lengthier periods under Prohibition shrinks over time, so late \( \delta_r \)'s are less precisely estimated. At its peak, cities were on average experiencing 3.1 homicides per 100,000 more than before Prohibition was introduced (s.e. = 2.7).

Analogous regression results for drunkenness arrests provide a complementary picture. The estimated \( \delta_r \)'s are presented in the right panel of figure 4. The figure illustrates the dramatic fall in drunkenness arrests during the first two years after a city was under Prohibition. This is the expected outcome of prohibiting the liquor trade, due to the impact closing of most of the supply sources of alcohol which, during this period, were to a large extent domestic. The reduction in the supply of alcohol is likely to be underestimated in figure 4, given that law enforcement does not show a fall during early Prohibition years, relative to years without Prohibition. During the second year under Prohibition, drunkenness arrests attain a minimum. The estimated coefficient for \( \delta_2 \) is \(-9.73\) (s.e. = 1.4), which implies that at its lowest point, the alcohol supply would have contracted 50\% \((= 12.7/19)\) in the absence of changes in law enforcement or demand. The figure also illustrates the steady recovery of the alcohol market, if we are willing to assume that arrest intensity did not change significantly throughout Prohibition. Approximately fifteen years into Prohibition, drunkenness arrests are indistinguishable from Pre-Prohibition levels\(^1\).

The patterns in panel A of figure 4 are consistent with the idea that legal Prohibition immediately had a large effect on the supply of alcohol. When looking at crime, it had a much smaller short-run impact, likely due to the slow development of alternative (illegal) sources of alcohol and their associated crime networks. On the other hand, the figure does not support the claim of Prohibitionists of the time, who claimed Prohibition would reduce criminality and the social disruptions associated with liquor consumption and the saloon; despite the large contraction of the alcohol market during the early prohibitionist years, a time when criminal organizations were still not developed, the homicide rate remained relatively steady.

Finally, panel B in figure 4 presents the estimates of the \( \delta_r \)'s both for the police share and for the

\(^1\)I call A the complete sample including all observations for which data is available, and also ran regression on it which I omit from the paper. Thoughout, no significant differences arise from results using either sample.

\(^{19}\)The identification assumption here is that the introduction of Prohibition did not also induce changes in individual's preferences over alcohol consumption. As an effort to check how reasonable this assumption is, Appendix 4 presents some evidence exploiting variation in the availability of neighboring alcohol markets. The evidence there is consistent with no changes in demand after the introduction of Prohibition.
per capita police expenditure. Both measures of law enforcement increase steadily until around
ten to twelve years into Prohibition, only to subsequently fall back at a mildly faster pace. The
pattern follows the one of the homicide rate; both variables appear to increase during the first years
of Prohibition, and to start falling at relatively similar times. Below I will argue that the rise and
fall patterns in police enforcement and crime can be understood as the equilibrium outcomes of a
dynamic learning process about the effects of Prohibition, and its interaction with the distribution
of moral preferences and the dynamics of the illegal alcohol market and its associated criminal
networks.  

4.2 Preferences and Moral heterogeneity

Communities with varying preferences over the legal standard were likely to collectively respond in
different ways to the introduction of Prohibition. The trends in figure 4 are likely to be averaging
out heterogeneous responses across cities with different moral profiles and beliefs, and thus, with
differing willingness to enforce the policy. On the one hand, drier constituencies should be willing
to enforce more because for the decisive voter, her marginal valuation of reducing her community’s
alcohol consumption was larger, and because she was likely to be more optimistic about the effects
of law enforcement under Prohibition. On the other hand, drier communities were likely to face
smaller potential alcohol markets, and hence less crime increases due to Prohibition. Thus, holding
moral views constant, the decisive voter in cities with larger drinking populations faced an incentive
to increase law enforcement, relative to cities with smaller alcohol markets. If moral views were
relatively fixed, changes in equilibrium law enforcement should be due to belief updating about the
effects of Prohibition.

The empirical analysis below is based upon comparing changes in outcomes over time, in cities having
different distributions of moral tastes, exploiting both the timing of adoption of Prohibition laws
and the variation in community preferences. I follow the historical literature, and use the variation
in the religious ascription distribution and in the ethnicity and age distribution of the population, as
the main observable characteristics correlated with moral views about alcohol Prohibition, and prior
beliefs about its effects. I construct a straightforward proxy for the “wet share” in the population,
$\mu_{ct}$, as the sum of the fractions of the population in any of the religions considered as “wet” in the
literature, the share of non-native white individuals, and the share of the population in the 15-44
years range. There is fairly widespread consensus that Baptist, Evangelical, Methodist, Mormon,
and Presbyterian religious ascriptions were more favorable to Prohibition, while Catholic, Orthodox,

---

20 Evidence that Prohibition enforcement was weakened after an “experimentation” also comes from the repeal of
enforcement laws in several states during the 1920s, as mentioned in section 2 when discussing the controversy over the
repeal of New York’s enforcement law. States under Prohibition before the adoption of the 18th Amendment had their
own alcohol enforcement legislation, which was in many cases strengthened or harmonized with federal legislation after
Congress passed the Volstead Act. All other states, with the exception of Maryland, adopted state-level enforcement
legislation right after the passage of the Volstead Act, thus complying with the shared-enforcement responsibilities
established by the 18th Amendment. Throughout the 1920s several states decided to repeal their state-enforcement
laws, effectively leaving the federal government alone in the enforcement of Prohibition. The state of New York took
the lead by repealing its enforcement law in 1923, very much against the will of the Federal government and of a large
share of upstate voters. It was followed by Montana in 1925, Nevada and Wisconsin in 1928, Massachusetts in 1930,
and Arizona, California, Colorado, Louisiana, Michigan, North Dakota, New Jersey, Oregon, and Washington in 1931.

21 I normalize this variable dividing by 3, the total measure of the religious, ethnicity, and age distributions.
Jewish, and Lutheran communities had much more positive views about alcohol consumption. On the other hand, while native whites, especially native white women, were strongly prohibitionist, foreign whites (Irish, Italians, Germans, Polish, Scandinavians) and blacks were more liberal about alcohol consumption. Finally, it is likely that younger populations also had more liberal views about liquor (See for example Foster (2002); Sinclair (n.d.); Szymansky (2003); Blocker (1989); Asbury (1950)). Thus I define “wetness” as:

$$\mu_{ct} = \frac{1}{3}(1 - \%Baptist_{ct} - \%Evangelical_{ct} - \%Methodist_{ct} - \%Mormon_{ct} - \%Presbyterian_{ct})$$

$$+ \frac{1}{3}(1 - \%NativeWhite_{ct}) + \frac{1}{3}(\%PopulationAges15 - 44_{ct})$$

In 1911, its mean is 0.49, with a standard deviation of 0.085. Similar to the empirical strategy in Equation (1), I regress each of the outcome variables $y_{ct}$ on the years-under-Prohibition indicators, and their interaction with the initial value of the “wetness” measure $^2$. As a benchmark for comparison, I ran analogous regressions using only the Constitutional Prohibition indicator, just as in the models in section 4.1. The models I estimate take the form:

$$y_{ct} = \alpha + \beta t + \sum_{r=1}^{k} \delta_{r}D_{cr} + \sum_{r=1}^{k} \phi_{r}D_{cr}\bar{\mu}_{c} + \gamma'X_{ct} + \varepsilon_{ct}$$

Interest lies in the differential evolution of outcomes over time under Prohibition, captured by the estimates of the $\phi_{r}$‘s, which measure how the several outcome variables changed differentially over the years under Prohibition, between cities with varying “moral” distributions (relative to a city with zero “wet” population). For ease of exposition, panel A in figure 5 graphs the estimated $\phi_{r}$‘s for the specifications using sample C. (See table A4-2, column (4), in Appendix 4). Estimates are very similar in magnitude for the alternative samples. The figure shows an increasing differential gap in the homicide rate during the first years under Prohibition, which subsequently closes over time, for cities with relatively “wetter” constituencies. This happened especially during the years in which the homicide rate was high. Because the differential increases in crime followed the same time pattern of overall crime during Prohibition, this suggests that a large fraction of the increase in criminality occurred in cities with wetter constituencies. Differential changes in the drunkenness arrest rate, which can be seen in the right panel of figure 5, appear to be small and significantly different from zero only in a few of the years under Prohibition when the alcohol supply was likely experiencing its fastest recovery.

Panel B plots the estimated $\phi_{r}$‘s for the police share and per capita police equations. Both show a similar pattern: cities with “wetter” constituencies increased police expenditure differentially less during early Prohibition years, but this gap closes over time, and for later Prohibition years, wetter

$^2$ I take 1911 as the baseline value for $\mu_{ct}$. For cities without religious distribution data before that year, I use the earliest year available (1916 in most cases). As a robustness check I ran identical regressions using the 1911 data on the somewhat reduced sample of cities without data before 1916, and results varied only marginally (available upon request).
cities have differentially higher spending in police. The relatively tighter law enforcement in drier cities during the early Prohibition years is consistent with their constituencies having relatively optimistic beliefs about its effects, making them more willing to repress the alcohol market, and expecting little response of crime. But criminality was increasing relatively more in wetter cities, and their alcohol markets were bouncing back faster. This suggests that criminality was very sensitive to the size of the potential alcohol market, requiring higher levels of crime enforcement in wetter cities, despite their preferences for a more lenient enforcement of the Prohibition laws. Indeed, panel B in figure 5 shows that changes in police expenditure were differentially higher in wetter cities during the later years under Prohibition. These were years in which cities were, overall, reducing police expenditure, so the figure implies that wet cities were unable to reduce law enforcement as fast. These patterns suggest that the alignment between the legal standard and community preferences played a major role in determining law enforcement outcomes. In cities where the median individual disfavored alcohol consumption and the alcohol market was small, there was little potential for crime to arise after the introduction of Prohibition; Prohibition enforcement could be tightened without concomitantly high crime increases. If individuals learn about the effects of Prohibition by observing crime outcomes, these communities should not alter their preferences too much over time. In contrast, communities where Prohibition was in stark contrast to average moral preferences over alcohol faced a much more demanding problem. In those cities, the alcohol market was large, so the potential for Prohibition-related criminality was much higher. The median citizen in this community should be unwilling to enforce Prohibition tightly, not only because she was likely to enjoy alcohol consumption and was morally liberal about others' alcohol consumption, but also because she was less optimistic about the response of criminality to Prohibition enforcement. This is what the early behavior of police expenditure suggests in figures 4 and 5, and is consistent with the repeal of state-level enforcement legislation.

If tightening Prohibition enforcement drove illegal producers towards a more intensive use of violence, why did police enforcement fall more slowly in wet cities in the later years under Prohibition, if these were the ones most unwilling to enforce it? I suggest the answer is the impossibility to separate overall crime enforcement and the enforcement of restrictions over a specific market, when the legal standard prescribes full Prohibition. The prohibited market itself becomes a major source of criminality, so that combatting crime also indirectly tightens the alcohol market. Under Prohibition, the ability to specifically target crime without restricting the alcohol market was limited, especially for policing activities. Thus, Prohibition in wet cities not only had adverse effects over crime, but also was costly because for a given level of police expenditure, it would lead to a larger response of crime relative to a city with a smaller alcohol market. This predicts larger shifts in preferences over Prohibition in these communities.

The timing of adoption of State-level Prohibition could be correlated with unobservables at the city level, which themselves would be causing the observed trends. However, this is unlikely because such trends should also have a non-linear effect over time. Moreover, I am looking at the effects of Prohibition on a sample of U.S. cities, which did not directly choose a Prohibitive legal standard.

---

23 As a placebo test for the results on police expenditure, I ran analogous models using the expenditure in fire. I do not include the results here to save space, but no discernible differences appear between cities with different moral profiles.
but rather saw it imposed upon them by state and federal decisions, making it less likely that
the timing of adoption of Prohibition is correlated with city-specific unobservables. Nevertheless,
other variation in previous legislation, in particular other alcohol-related laws, and women’s suffrage,
appear as potential correlates of the introduction of Prohibition. In Appendix 4 I look at variation in
the availability of neighboring alcohol supply sources, in pre-Prohibition state-level alcohol-related
legislation, and in women’s suffrage legislation. The evidence does not suggest that alternative
legislation was driving the patterns described above.

4.3 Public Opinion

To look at changes in political support for Prohibition during this period I exploit alcohol-related
referenda election returns, available at the county level for most of the U.S. states, taking place in
different years during the 1910s-1930s. I focused on finding for each state, electoral returns on a liquor
referendum taking place prior to the introduction of Prohibition in the State (the pre-Prohibition
period), and for a year in the later Prohibition period or after the repeal of federal Prohibition (the
post-Prohibition period). Because most of the information is available at the county level, here I
present results for both a county panel and a city panel, assigning the county vote to the city(ies)
in the county. A comparison of the distribution of wet vote shares prior to and after Prohibition
reveals the dramatic shift in public opinion. Figure 6 presents the histograms of county wet vote
shares in both periods. In the pre-Prohibition referenda, the 75th percentile of the distribution
of wet vote shares is 0.5. Thus, in three quarters of the counties some type of Prohibition had
majoritarian support. In the post period, only 35% of counties had majorities favoring Prohibition.
On the other hand, the comparison of both histograms suggests a spreadout in the distribution of
public opinion regarding the policy.

Figure 7 shows differential patterns of opinion shift between communities with varying moral profiles.
There was strong convergence of public opinion against Prohibition, but it was restricted to commu-
nities that were morally more favorable to alcohol in the first place. The figure breaks the sample of
counties between those with a value of my “moral wetness” measure, $\mu$, below (left figure) and above
(right figure) the median of 0.355, and plots the pre-Prohibition and the post-prohibition wet vote
shares in the horizontal and vertical axes respectively, together with a 45 degree line. Almost all
counties above the median had a public opinion shift against Prohibition, while in the set of below
median counties, a considerable fraction even observed shifts towards Prohibition. Moreover, among
the latter group of counties there is no evidence of “convergence of opinion”, since pre-Prohibition
vote shares are a very good predictor of post-prohibition ones. In contrast, among above-median
counties the shift against Prohibition was on average much larger in counties initially more favorable
to Prohibition. The shift in public opinion was concentrated in the upper part of the distribution
of moral preferences. Analogous figures for the city sample reveal the same patterns.

More formally, I estimate fixed-effects regressions for both the county and the city samples, with two

\[^{24}\text{Except for cities in Massachusetts and Connecticut, for which city-level data is available.}\]
\[^{25}\text{I computed } \mu \text{ for each county directly from equation 2 for county-level data. I used the 1916 and 1926 Census of}\]
\[^{26}\text{Religions for the religious ascriptions distribution. For the age and ethnicity distributions I used the 1920 and 1930}\]
\[^{27}\text{Population Censuses because the county-level age distribution from the 1910 census is unavailable.}\]
periods, \( t \in \{0, 1\} \). \( t = 0 \) is the pre-Prohibition period, and \( t = 1 \) is the post-Prohibition period, for a year in which there was a liquor-related referendum. The models I estimate take the basic form

\[
w_{ct} = \alpha_c + \beta t + \delta \mu_{ct} + \phi \mu_{ct} t + \gamma' X_{ct} + \varepsilon_{ct}
\]

where \( w_{ct} \) is the wet vote share. In this model the interaction term for the post period uses the initial period’s wetness, given that it is based on baseline moral preferences that law enforcement and its equilibrium effects are endogenously determined. \( X_{ct} \) is a vector of time-varying controls, including the log of population (1910 data for \( t = 0 \) and 1930 for \( t = 1 \)), the urban share of the county (or of the county’s city), the number of dry laws in place, the year in which the referendum took place, and indicator variables for the type of referendum (a Prohibition law, a constitutional convention election or a constitutional amendment (omitted category)). The estimate of \( \phi \) should capture the differential increase in the wet vote share in wetter communities.

Table 3 presents the main results. Columns (1) – (5) present results for the complete sample of counties. For comparative purposes, columns (6) – (10) present estimates for analogous models but restricting the sample to counties with a population larger than 30,000. Finally columns (11) – (15) present results for the sample of cities. Columns (1), (6), and (11) first simply regress the wet vote share on a post-Prohibition period indicator. The estimated coefficient in column (1) implies that the average county experienced a 13 percentage points larger wet vote share after Prohibition (s.e. = 0.004). Column (2) then presents estimates of the main specification in equation (4) without additional controls. Column (3) controls for the log of population and the urban share, the year in which the referendum took place, and indicators for the type of referendum. Both the type of referendum in consideration and the year in which it took place are likely to be endogenous to the vote share, given that the timing and kind of referendum were likely to depend on the trends of public support for Prohibition; for example, a proposal for a constitutional amendment was likely to take place in states where public opinion favoring Prohibition was believed to be widespread. Thus, I do not stress the results of the models in columns (3), (8), and (13); nonetheless, estimates are very similar to those excluding these variables. Column (4) includes state-cross-post Prohibition interactions, and finally column (5) accounts for the potential selection problem arising from the fact that a subset of wet states never held pre-Prohibition liquor referenda, by controlling for the inverse Mills ratio of the estimates of a Probit selection equation for holding a referendum (See Appendix 4). In all regressions I run a completely balanced panel. The estimates of the selection equation are shown in panel B. If anything, the size of \( \phi \), the estimated differential effect of having a larger wet constituency, increases when accounting for selection.

The estimate of \( \phi \) from column (5) implies that a county with a one standard deviation higher \( \mu_{ct} \) would differentially increase its wet vote share by 6 percentage points (0.062 = 0.48 x 0.13). The interaction terms are very precisely estimated across specifications, and the regression results suggest that most of the increase in support for anti-Prohibitionism occurred through the differentially larger growth in wet support of morally wet communities. The magnitude and significance of the estimates for the city sample are very close to those of the county sample, as can be seen in columns (11) – (15). Estimates for the restricted sample of more populous counties are even larger in magnitude, and imply that the result is not driven by a comparison of extremely dry versus extremely wet communities.
Given that wetter communities were initially less in favor of Prohibition, there was less room for an increase in anti-Prohibitionist sentiment. Nevertheless, the referenda electoral results suggest these communities did experience larger public opinion shifts. One possibility is that learning in dry communities was slower because of large differences in prior beliefs about the effects of the policy, coupled with uninformative local law enforcement decisions. It could also be that all communities were learning at similar speed, but that differences in moral views were so large that for the driest of communities indirect preferences over the policy were very inelastic to changes in beliefs. Finally, it is also possible that very dry communities did in fact benefit from Prohibition. Although this seems at odds with the reduced-form results on crime presented above, the experience of rural and very dry counties might have been very different, given that local preferences were much more aligned to a prohibitionist legal standard.

Overall the reduced-form results show that the introduction of Prohibition had heterogeneous effects across cities varying in their moral preferences over the policy, and directly point towards a set of elements that a comprehensive theory of endogenous law enforcement in the context of U.S. Prohibition should incorporate. First, that responses varied over time, and that restricting attention to Nationwide Prohibition is insufficient to understand the trends in the different outcomes focused on; the passage of state legislation and enforcement laws, together with local law enforcement decisions, appear as first order. Second, that learning about the effects of the policy is likely to have driven not only the evident changes in public opinion but also the equilibrium law enforcement choices during Prohibition years. Third, that the dynamics of the alcohol market were important for the evolution of criminality during Prohibition. Finally, that variation in the potential alcohol markets across cities implied differential constraints on the extent to which communities could vary law enforcement.

5 A Statistical Model of Prohibition, Learning, and Endogenous Law Enforcement

In this section I develop a simple political economy model of Prohibition enforcement and learning. It incorporates the central interactions at the heart of the dynamics of criminality and public opinion during Prohibition, based on the discussion above. It provides enough structure to be directly estimated. Importantly, it is an equilibrium dynamic model where equilibrium outcomes are the result of the optimal choices of agents, and where learning is rational. Prohibition altered the Data Generating Process (DGP) of several economic outcomes through two main channels. First, in the absence of Prohibition there is no direct link between law enforcement and criminality; this link arises through the enforcement of dry legislation when Prohibition is adopted. Second, differences in beliefs and uncertainty about the effects of Prohibition created a new dynamic channel affecting law enforcement choices at the local level, because many communities were experimenting a new legal standard with unknown consequences at the time of its adoption. In the model, the interaction between moral preferences and beliefs determines the political-equilibrium choices of law enforcement which, by affecting crime, determines the endogenous evolution of learning about the effects of the policy. The evolution of beliefs subsequently shift optimal law enforcement choices and public opinion.
over Prohibition. This requires that individuals know the mapping from indirect preferences to law enforcement choices (the political process), and have beliefs about the mapping from law enforcement to expected outcomes.

5.1 Environment and Preferences

Consider a society made up of a large number of small communities \( c = 1, 2, \ldots \), in discrete time. Community \( c \) is populated by a continuum measure \( 1 \) of adult citizens indexed by \( i \). Each period \( t = 0, 1, 2, \ldots \), every citizen makes a private decision about alcohol consumption, and through majority voting, collectively decides how to distribute a fixed public budget among public goods. Each adult lives for one period, and has a child\(^{26}\).

In addition, society as a whole (out of which the individual community is small) can decide a legal standard over the alcohol market for the community, either to be under Prohibition \((P_t = 1)\) or not under Prohibition \((P_t = 0)\). In the latter regime, alcohol is freely traded (though possibly with some regulation), whereas in the former, an illegal alcohol market is the only source of liquor. Under no Prohibition the alcohol market is perfectly competitive, while under Prohibition, the black market is monopolistic. When Prohibition is in place, the community collectively decides the extent of enforcement of the law. Finally, \( P_0 = 0 \), so that society’s initial legal standard is liberal.

Citizens are heterogeneous in several private and common-values dimensions (Arrow (1963)). In regard to private values, each adult citizen is either dry \( D_t \) or wet \( W_t \), and \( I \) denote \( \mu_t = |W_t| \) as the share of wet adult citizens. The two groups differ in their preferences over individual alcohol consumption \( h \). For simplicity, dry individuals do not derive any utility from their own consumption of alcohol, while wet adult individuals do enjoy consuming a unit of alcohol every period \((h \in \{0, 1\})\). This type is not inherited from parent to child, but during every period the share of wet individuals is a random variable drawn from a beta distribution (See Coate and Conlin (2004) or Degan and Merlo (2009) for a modeling choice in the same spirit):

\[
f_\mu(\mu; a, b) = \frac{\mu^{a-1}(1 - \mu)^{b-1}}{\int_0^1 v^{a-1}(1 - v)^{b-1}dv}, \quad a, b > 0
\]

Individuals know the parameters of the distribution, but do not observe the draw directly (they do not observe the type of their fellow citizens). Each individual is also characterized by a “moral view” \( z^i \), which is a measure of the marginal disutility she gets from her community-wide alcohol consumption. I will call \( z^i \) her moral view, and will assume it is inherited from parent to child.

On the other hand, individuals in the community have common values about consumption of a public good \( G \), and crime, but there is heterogeneity in prior beliefs (beliefs of the cohorts living prior to and during the first period under Prohibition) about how the introduction of Prohibition might impact crime within the community. Thus, conflicting views over Prohibition arise not only from differences in individual moral stands (tastes), but also from informational differences (or differences in the way prior information was interpreted). Nevertheless, these are correlated in the population.

\(^{26}\)Throughout this section I drop the community indices \( c \), since no confusion arises. In section 6.1 specify which parameters are city-specific for estimation purposes.
to allow individuals with more radical views against alcohol consumption (by others) to be more optimistic about the response of criminality to Prohibition.

Specifically, the information structure, which will imply parsimonious learning dynamics, is as follows. Individual $i$’s moral view (distaste for her community’s aggregate alcohol consumption) is $z^i = z + \zeta^i$, where $z$ is her community’s average moral view, and $\zeta^i$ is her individual-specific moral shock. On the other hand, her prior beliefs (about the elasticity of crime to the enforcement of Prohibition, as will be explained below) are $\theta_0^i = B + \xi^i$, where $B$ can be thought of as the common component of prior beliefs (which possibly includes a bias), and $\xi^i$ is an individual-specific bias. $(\zeta^i, \xi^i)$ is drawn from a joint-normal distribution

$$\begin{pmatrix} \zeta^i \\ \xi^i \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_\zeta & \rho \sigma_\zeta \sigma_\xi \\ \rho \sigma_\zeta \sigma_\xi & \sigma^2_\xi \end{pmatrix} \right)$$

(6)

Here moral views are understood as the set of beliefs about the world, which an individual takes as true. This is, to which she assigns a degenerate prior probability of 1, and are thus not subject to updating with the arrival of new information. On the other hand, all other beliefs can evolve through rational updating as the individual receives new information. In the context of Prohibition, it is natural to think of crime as the source of information about $\theta$. Observe that if $\rho < 0$, individuals who have stronger moral views against alcohol will be on average more optimistic about the response of crime to the introduction of Prohibition. For simplicity, both wet and dry individuals get their $(\zeta^i, \xi^i)$ drawn from the same distribution.

The expected utility of a citizen is given by

$$E_i U^i(h_t^i, A_t, G_t, q_t | P_t) = E \left[ 1_{\{i \in W_t\}} h_t^i - z^i A_t + V(G_t) - q_t \right]$$

(7)

where $A_t$ is the aggregate alcohol consumed in his community, $q_t$ is the crime rate, $G_t \in [0, 1]$ is the share of the public budget allocated to public goods other than policing, and $E$ is the expectations operator conditional on all the information available to individual $i$. The term $-z^i A_t$ represents the “moral externality”. Finally, $V(G) = \exp(G)$. Notice that from the point of view of individuals the optimization problem is static, since they only live for one period.

5.2 The Alcohol Market

Imagine a very simple alcohol market, where the price of consuming a unit of liquor is normalized to zero under no Prohibition, but individuals must engage in a costly search. The probability of successful search is a decreasing function of the level of Prohibition enforcement chosen by the community. As communities decide to tighten enforcement of dry laws, the availability of alcohol is diminished. Specifically I allow this probability to take the form $Pr(h_t^i = 1 | P_t = 0) = \exp(-e_t)$ where $e_t \geq 0$ is the level of dry law enforcement.

\[27\] Recall that under no Prohibition dry laws were in place. These restricted the availability of liquor by regulating the alcohol market along different dimensions.
The introduction of Prohibition, on the other hand, makes legal alcohol unavailable (and increases market power since black markets are likely to be captured by a small set of criminal organizations). The search for alcohol becomes costlier, and I will allow the probability of a successful search to also become a function of the amount of time the community has been under Prohibition, \( \tau_t \), to flexibly capture the possibility that the illegal market adjusts over time. After Prohibition is adopted, the legal market for alcohol is closed on impact, which by itself has an effect on the quantities traded. The supply response from illegal producers does not occur immediately because it takes time to build up a black market, and the development of crime networks associated with the illegal activity also requires costly and staggered investments. Thus, the probability of successful search under Prohibition is given by \( \Pr(h^1_t = 1|P_t = 1) = k(\tau_t)\exp(-e_t) \), where

\[
k(\tau_t) = 1 - \lambda \tau_t \exp(-\kappa \tau_t)
\]

with \( \kappa, \lambda > 0 \). It follows that aggregate alcohol consumption is

\[
A_t(e_t) = \int_{\tau \in W_t} 1k(\tau_t)\exp(-e_t)\,d\tau = \mu_t k(\tau_t)\exp(-e_t)
\]

so that during the \( \tau \)th year under Prohibition, holding law enforcement constant, the alcohol market is a fraction \( k(\tau_t) \) of what it would be under no Prohibition. This highlights why individuals with moral views opposed to alcohol might want to choose high levels of law enforcement. By reducing the equilibrium consumption of alcohol, their moral externality is directly reduced. The fact that after an initial fall \( k(\tau_t) \) rises as time under Prohibition increases, implies that over time, higher levels of law enforcement are required to maintain a given size of the illegal alcohol market.

\[5.3 \text{ Crime, Prohibition, and Law Enforcement}\]

I allow crime to be related to alcohol consumption by assuming that baseline crime is proportional to the size of the alcohol market. Formally,

\[
q^N_t = \Theta_S + A(e_t) = \Theta_S + k(\tau_t)\mu_t\exp(-e_t) + \varepsilon_t
\]

where \( \varepsilon_t \sim N(0, \sigma^2_\varepsilon) \) is an iid normally distributed shock. Because the homicide rate levels vary significantly across states but are relatively similar between cities in the same state, I allow for a state-specific parameter \( \Theta_S \). Central to the understanding of the variation in criminality across the United States during Prohibition is the fact that different communities were structurally different in how the ban on the alcohol trade would affect criminality, and there was disagreement about...

---

28 I introduce two parameters for \( k(\tau_t) \) to be flexible enough to separately capture the initial fall in the alcohol market once Prohibition is enacted (\( \lambda \)), and the speed at which the alcohol market bounces back (\( \kappa \)), and will restrict them to be constant across cities in the empirical analysis below. Note that for no-Prohibition years, \( k(\tau_t) = k(0) = 1 \). A graph of \( k(\tau_t) \) is presented in the first panel of figure 11 for \( \kappa = 0.26 \) and \( \lambda = 0.25 \) (the MLE estimates). This curve has its unique minimum at \( \tau_t = 1/\kappa \), so that \( \kappa \) is also the inverse of the time at which the alcohol market reaches its maximum size. I also impose the condition \( \kappa \exp(1) > \lambda \), which is necessary and sufficient for \( k(\tau_t) \) to be everywhere positive. A comparison of figures 2 and 11 illustrates why the functional form choice in 8 is likely to be appropriate.
this issue. I will assume the following relationship between law enforcement and Prohibition-related crime:

\[ q_t^P = \theta [A_t(e_t = 0) - A_t(e_t)] = \theta k(\tau_t) \mu_t [1 - \exp(-e_t)] \]  

Equations (10) and (11) capture the two main channels from the alcohol market to crime. Alcohol consumption can cause crime by altering the behavior of consumers, and by giving incentives for the development of crime networks when it is prohibited \(^{29}\). Total crime is \( q_t = q_t^N + P q_t^P \). In equation (11), \( \theta \) is the true state, a city-specific shifter of crime to the size of the alcohol market under Prohibition. Formally, this implies a structural change in the Data Generating Process for crime when Prohibition is introduced. For \( \theta > 0 \), it measures the extent to which crime increases as the alcohol market is tightened through law enforcement, relative to the size of the market at zero law enforcement. Observe that \( q_t^P = 0 \) if \( e_t = 0 \), or under no Prohibition. Also, as \( e_t \to \infty \), Prohibition-related crime \( q_t^P \to \theta e_t k(\tau_t) \mu_t \). This functional form captures a set of key aspects about the link between criminality and law enforcement under Prohibition. First, sustaining a smaller black market when alcohol is prohibited, translates into more crime. Second, a larger wet share implies a larger potential alcohol market, and hence, more Prohibition-related crime for a given level of law enforcement. Third, the time-variation in crime should be correlated with the time-variation in the alcohol-market dynamics. Fourth, and most importantly, a link between restrictions in the alcohol market and criminality only appears when Prohibition is in place. There is common knowledge up to the uncertainty about the value of \( \theta \).

The drunkenness arrest rate is, by definition, the conditional probability of being arrested times the alcohol market size. It is a function of law enforcement, and I will allow the probability of being arrested to take the flexible form \( Pr(Arrest|e_t) = \frac{\exp(e_t)}{\chi + \exp(e_t)} \), with \( \chi > 0 \). The drunkenness arrest rate is thus:

\[ d_t = Pr(Arrest|e_t) A_t(e_t) = \frac{\mu_t k(\tau_t)}{\chi + \exp(e_t)} \]  

Notice this equation holds both under no Prohibition and under Prohibition, since under no Prohibition public drunkenness was also prosecuted. The equilibrium drunkenness arrest rate is a decreasing

\(^{29}\)In a classic Sociology paper, Paul Goldstein discusses the different channels from drug use to violence. The author identifies two sources of criminality in a no Prohibition environment: psychopharmacological and economically compulsive: In the former, "... some individuals, as a result of short or long term ingestion of specific substances, may become excitable, irrational, and may exhibit violent behavior". In the latter, "...some drug users engage in economically oriented violent crime, e.g., robbery, in order to support costly drug use.... Violence generally results from some factor in the social context in which the economic crime is perpetrated." Then he identifies systemic violence as an added source of crime under Prohibition: "... the aggressive patterns of interaction within the system of drug distribution and use... 1. disputes over territory between rival drug dealers. 2. assaults and homicides committed within dealing hierarchies as a means of enforcing normative codes. 3. robberies of drug dealers and the usually violent retaliation by the dealer or his/her bosses. 4. elimination of informers. 5. punishment for selling adulterated or phony drugs. 6. punishment for failing to pay one’s debts. 7. disputes over drugs or drug paraphernalia. 8. robbery violence related to the social ecology of coping areas." Goldstein (1985, pp.146-149)

\(^{30}\)The choice of this logistic functional form for the conditional probability of being arrested under drunkenness charges is flexible enough to allow any arrest probability at zero law enforcement: \( Pr(Arrest(0) = 1/(1 + \chi) \), which is a convenient way to interpret \( \chi \).
function of law enforcement. Equation (12) highlights that variation in the drunkenness arrest rate can come from changes in the size of the alcohol market, (the wet share \( \mu_t \) and the “secular” dynamics of the alcohol supply under Prohibition \( k(\tau_t) \)), or from the extent of law enforcement \( e_t \). Moreover, when identifying these two channels separately, the structural estimation will exploit the common variation in drunkenness arrests, crime, and police expenditure due to changes in the size of the alcohol market and in law enforcement.

Prohibition enforcement is a function of the amount of police expenditure \( p_t \), and the current legal standard, which includes dry laws, enforcement laws, and Prohibition. I will assume Prohibition enforcement can be expressed as \( e_t = \alpha_t p_t \), with \( \alpha_t > 1 \), which depends on the legal standard in place. The multiplicative form is intended to capture the inherent non-separability between crime and Prohibition enforcement. Observe, nonetheless, that liberalizing the legal standard (by lowering \( \alpha_t \)) weakens the link between both, at the cost of reducing the restrictions on the alcohol market. Each community has a unit of public resources to allocate between policing \( p_t \) and other public goods \( G_t \), and I assume, for simplicity, they can be exchanged one-for-one. Thus,

\[
G_t = 1 - p_t
\]

5.4 Learning and the Timing of Events

To make the model suitable for estimation, I make the following assumptions about information, learning, and the timing of events. In the end of period \( t - 1 \), each member of the adult cohort has one child, and outcome variables \((p_{t-1}, q_{t-1}, d_{t-1})\) are realized. Under no Prohibition there is no learning taking place, whereas in a Prohibition year, children observe the vector of outcome variables and update their beliefs about \( \theta \) according to Bayes’ rule. This occurs as follows. First, each child learns her parent’s belief \( \theta^c_{t-1} \). In the first year under Prohibition \((\tau_{t-1} = 1)\), child \( i \) knows that \( \theta^0 = B + \xi^i \) (of course, she does not observe \( B \) or \( \xi^i \) separately), and knows that \( \xi^i \sim N(0, \sigma^2) \) is the marginal distribution of biases in the population. As a result, child \( i \)'s prior about \( \theta \) is given by \( \theta^i_{t-1} \sim N(\theta^0, \sigma^2) \).

From equation (12), after the child has observed \( d_{t-1} \) and \( p_{t-1} \), she can perfectly back-up the realization of \( \mu_{t-1} \). Thus, in the public signal \( q_{t-1} = \Theta_S + \mu_{t-1} k(\tau_{t-1}) \exp(-\alpha_t p_{t-1}) + \theta k(\tau_{t-1}) \mu_{t-1} [1 - \exp(-\alpha_t p_{t-1})] + \varepsilon_{t-1} \), the only remaining uncertainty comprises the true value of \( \theta \) and the distribution of \( \varepsilon_{t-1} \). It follows that Bayesian individuals’ posteriors about \( \theta \) will be normally distributed. Normal updating will keep taking place cohort after cohort as long as the community is still under Prohibition. Thus, iteratively using normal updating and exploiting linearity of conditional distributions under normality, cohort \( t \)'s posterior (or \( t + 1 \)'s prior) will be given by

\[
\theta^c_t \sim N \left( \frac{1}{\frac{1}{\sigma^2_t} + \frac{1}{\sigma^2_q} \sum_{s=m}^{t-1} \omega_s^2} \theta^c_{t-1} + \frac{1}{\sigma^2_t} \sum_{s=m}^{t-1} \omega_s^2 \sum_{s=m}^{t-1} \frac{1}{\sigma^2_q} \exp(-\alpha_t p_{t-1}) \omega_s, \frac{1}{\sigma^2_t} + \frac{1}{\sigma^2_q} \sum_{s=m}^{t-1} \omega_s^2 \right)
\]

where \( s_0 \) is the first year in which community \( c \) is under Prohibition, and where \( \omega_t \equiv k(\tau_t) \mu_t [1 - \exp(-\alpha_t p_{t-1})] \).
$\exp(-\alpha pt)$ is a measure of the degree of informativeness of the signal \(^{31}\). This posterior will be the relevant measure with respect to which individual \(i\) will evaluate her expected utility under different law enforcement policy alternatives.

The stochastic process in (14) is a bounded martingale, and as such, the \(\{\theta_t^i\}\) converge almost surely as \(t \to \infty\). Moreover, because the true distribution (a mass point of 1 at \(\theta\)) is absolutely continuous with respect to the prior (which is normal and hence has positive density everywhere), the process will converge to the true \(\theta\) for any infinite sequence of positive \(\{\mu_t, \epsilon_t\}_{t=0}^{\infty}\). How rapidly convergence occurs will depend on the amount of law enforcement. As \(p_t \to 0\), the signal becomes uninformative because individuals know the data generating process, and hence, realize that at zero enforcement any observed crime rate must not come from Prohibition-related crime. Conversely, for a given observed signal, a higher value of law enforcement reduces the variance of the signal’s likelihood, making its informational content much higher. Rational individuals should then put a higher weight on such a signal. Interestingly, this implies that if a community reduces its enforcement levels, it will also reduce the speed at which its members will be able to learn about the true state. Now I define \(\theta_t^i\) as the posterior mean, and express it more compactly as

\[
\theta_t^i = \Omega_t \frac{1}{\sigma_x^2} \theta_0^i + \Omega_t \frac{1}{\sigma_z^2} \sum_{s=0}^{t-1} [q_s - \Theta_S - \mu_s k(\tau_t) \exp(-\alpha t p_s)] \omega_s = \Omega_t \frac{1}{\sigma_x^2} \xi^i_t + \Omega_t \theta_t^C \tag{15}
\]

where \(\Omega_t \equiv \frac{1}{\sigma_x^2 + \frac{1}{\sigma_z^2} \sum_{s=0}^{t-1} \omega_s^2}\) is the posterior variance, and the common component of beliefs (shared by all individuals in the community) is \(\theta_t^C \equiv \frac{1}{\sigma_x^2} B + \frac{1}{\sigma_z^2} \sum_{s=0}^{t-1} [q_s - \Theta_S - \mu_s k(\tau_t) \exp(-\alpha t p_s)] \omega_s\). The posterior mean belief at any time \(t\) is a weighted average of the prior mean and the whole history of crime realizations, weighted according to their relative precisions and by the informativeness of each signal. The degree of informativeness depends, in turn, on the extent of law enforcement originating the signal. Equation (15) shows that individual belief sequences can be analytically decomposed into a common component, shared by all individuals in the community every period given the public nature of the signal, and an individual-specific component, tied to the dynasty-specific bias. Of course, individuals do not separately observe the public and the private components of their beliefs, but the explicit distinction will be convenient. Equation (15) is readily interpretable. When the precision of the distribution of prior biases is low (as measured by \(1/\sigma_x^2\)), Bayesian individuals will disregard the information in their prior and will rely more closely on the observed signal sequence. A lower precision of the signal (\(1/\sigma_z^2\)) induces a Bayesian individual to put more weight on her prior. Moreover, since individuals know the DGP up to the uncertainty about \(\theta\), they optimally use the information on law enforcement to decide how much weight to give to the crime signal.

\(^{31}\)Equation (14) above highlights the convenience of assuming normality for both the prior on \(\theta\) and the conditional likelihood of the signal which, being an affine information structure, results in a very parsimonious learning process where posterior conditional expectations are linear in the signal sequence, making estimation relatively straightforward. Although this seems to be a very restrictive set of assumptions about the information structure and of the cognitive requirements of the learning process, these features of normal learning are actually fairly robust to alternative specifications. For example, if agents are not fully Bayesians, and are limited to making the best linear predictions based on the signal sequence \(\{q_t\}_{t=0}^{t-1}\), their prediction of the conditional mean will exactly match the posterior mean under normal updating, no matter the true data generating process (See for example, Vives (2010, p. 379)).
5.5 Political Equilibrium and the Distribution of Preferences over Law Enforcement

5.5.1 The Problem under no Prohibition

Replacing the probability of successful search, and equations (9), (10), and (13) into (7), indirect preferences under no Prohibition can be obtained. The first order condition implies that the preferred police enforcement of individual \( i \) is given by (see Appendix 1),

\[
\frac{1}{\alpha_t - 1} \left\{ \ln \alpha_t + \ln \left[ \frac{a}{a + b} (z + 1) + \frac{a}{a + b} \zeta^i - 1_{\{i \in W_t\}} \right] - 1 \right\} = p_t^i(\zeta^i) = \frac{1}{\alpha_t - 1} \left\{ \ln \alpha_t + \ln \left[ \frac{a}{a + b} (z + 1) + \frac{a}{a + b} \zeta^i - 1_{\{i \in W_t\}} \right] - 1 \right\}
\]

(16)

If the expression inside \( \ln[] \) is negative, \( p_t^i(\zeta^i) = 0 \). This expression follows from the fact that \( \mu_t \) is distributed \( \beta(a, b) \), so its mean is given by \( \frac{a}{a+b} \), and that the expected alcohol consumption for a wet individual is equal to the probability of successful search. When a community is not under Prohibition, beliefs about \( \theta \) do not appear in the objective function of its members. The ideal choice of police enforcement simply trades off the reduction in other public goods with the reduction in moral externality from tightening the alcohol market, and the reduction in overall crime. Individuals with higher \( z^i \) will prefer higher levels of law enforcement.

Equation (16) illustrates clearly some of the interesting interactions in the context of moral conflict. Wet individuals, who suffer a small moral externality from average alcohol consumption, prefer low levels of policing to reduce the size of the market, but differentially higher the larger is the alcohol market in their community (the larger is \( a/(a + b) \)). Interestingly, this interaction effect is not present for dry individuals; for them, the marginal disutility of a larger alcohol market induced by a reduction in policing is exactly offset by the marginal disutility of increased criminality brought about by such a reduction in crime enforcement. The effect of tightening the legal standard on the ideal choice of policing, on the other hand, is ambiguous, since it trades off the value of reducing expenditure in police with the complementarity of police enforcement and the legal standard. For large values of \( \alpha_t \) though, ideal policing is falling in \( \alpha_t \).

5.5.2 The Problem under Prohibition

Taking a look at the problem under Prohibition by replacing the successful-search probability and equations (9), (10), (11) and (13) into (7), indirect preferences under Prohibition are obtained. From the first order condition, the preferred police enforcement of individual \( i \) under Prohibition is given by (see Appendix 1),

\[
\frac{1}{\alpha_t - 1} \left\{ \ln \left[ \alpha_t k(\tau_i) \right] + \ln \left[ \frac{a}{a + b} \left( z - \Omega_t \bar{d}^{C} + 1 \right) + \frac{a}{a + b} \left( \zeta^i - \Omega_t \frac{1}{a \xi} \xi^i - 1_{\{i \in W_t\}} \right) - 1 \right] - 1 \right\} = p_t^i(\zeta^i, \xi^i) = \frac{1}{\alpha_t - 1} \left\{ \ln \left[ \alpha_t k(\tau_i) \right] + \ln \left[ \frac{a}{a + b} \left( z - \Omega_t \bar{d}^{C} + 1 \right) + \frac{a}{a + b} \left( \zeta^i - \Omega_t \frac{1}{a \xi} \xi^i - 1_{\{i \in W_t\}} \right) - 1 \right] - 1 \right\}
\]

(17)

where I have made use of equation (15). Once again, if the expression inside \( \ln[] \) is negative, \( p_t^i(\zeta^i, \xi^i) = 0 \) is the preferred police enforcement share. What matters for individual \( i \) is his mean
belief about $\theta$. Under Prohibition, individuals must now include the increased criminality induced by Prohibition enforcement in their optimal trade-off regarding police expenditure. Equation (17) highlights that the introduction of Prohibition alters individuals' optimal degree of law enforcement, which now becomes a function not only of their wet or dry identity and their dynasty-specific moral shock $\zeta^i$, but also of their dynasty-specific belief bias $\xi^i$. These are the three sources of unobserved heterogeneity in the model.

The analysis above looked at the indirect preferences of individuals over law enforcement. Nevertheless, law enforcement is a collective decision, which is made through majority voting. Thus, I define an equilibrium of this model as follows:

**Definition.** An equilibrium is a sequence of police expenditure shares $\{p^*_t\}_{t=0}^{\infty}$ such that for every $t$, $p^*_t$ wins any pairwise vote against any other $p^t_i$ when all adult citizens vote sincerely given their current beliefs $F^i_t(\theta)$, sequences of homicide and drunkenness arrest rates $\{q^i_t\}_{t=0}^{\infty}$, $\{d^i_t\}_{t=0}^{\infty}$ given by (11) and (12), and a sequence of belief distributions $\{F^i_t(\theta)\}_{t=0}^{\infty}$ for each dynasty $i$, which are updated every period according to Bayes' rule and given by (14).

To find the equilibrium path, it is necessary to look at the collective decision-making process, which takes the form of simple majority voting. Although there are three sources of heterogeneity regarding preferences over law enforcement across individuals in this model, below I show they can be reduced to one dimension, over which a unique majority-voting equilibrium exists.

**Proposition 1.** For any $t$, a given distribution of beliefs $F^i_t(\theta)$ $\forall i \in [0,1]$, and a legal standard vector $(\alpha_t, \tau_t, P_t)$, there is a unique equilibrium level of law enforcement $p^*_t$ given by

$$p^*_t = \frac{1}{\alpha_t - 1} \left\{ \ln \left[ \alpha_t k(\tau_t) \right] + \ln \left[ \frac{a}{a + b} \left( z - P_t \Omega_t \tilde{g}^C_t + 1 \right) + (1 - P_t) \tilde{g}_N^\text{med} + P_t \tilde{g}_P^\text{med} \right] \right\} - 1 \right\}$$

(18)

where $\tilde{g}_N^\text{med}$ and $\tilde{g}_P^\text{med}$ are random variables whose densities $f_{\tilde{g}_N^\text{med}}(\tilde{g}_N^\text{med}; a, b, \sigma_{\tilde{g}_N})$ and $f_{\tilde{g}_P^\text{med}}(\tilde{g}_P^\text{med}; a_c, b, \sigma_{\tilde{g}_P})$ are continuous and positive over the interval $[-1,0]$.

**Proof.** See Appendix 2. □

As the proof of Proposition 1 shows, $\tilde{g}_N^j$ and $\tilde{g}_P^j$ are one-dimensional sufficient statistics capturing the three sources of heterogeneity in individual $i$'s preferences, during no Prohibition and Prohibition periods, respectively. Their conditional distribution across the population is a mixture of two normal densities, weighted by the wet share $\mu_t$. In Appendix 2 I show that the equilibrium level of law enforcement is determined by the median voter's value of $\tilde{g}_j^i$. Because the wet share is itself a beta-distributed random variable, $\tilde{g}_j^\text{med}$ is also a random variable whose density $f_{\tilde{g}_j^\text{med}}(\tilde{g}_j^\text{med}; a, b, \sigma_{\tilde{g}_j})$ is continuous and takes positive values over the interval $[-1,0]$. As $\mu_t \to 1$, $\tilde{g}_j^\text{med} \to -1$, and as $\mu_t \to 0$, $\tilde{g}_j^\text{med} \to 0$. When all the community is wet, for example, $\mu_t = 1$ so the median in the community corresponds to the median over the distribution of preferences of wet individuals. These are normally distributed with mean and median at $-1$, given the preference for private alcohol consumption of wets.
5.6 Predictions and Main Assumptions

5.6.1 Predictions

The model makes several predictions about the equilibrium dynamics of law enforcement during Prohibition. First, observe the dynamic trade-off faced by a relatively liberal median voter living in a relatively wet community. Her pessimism about Prohibition-related crime ($\theta_t > 0$) makes her prefer a low level of law enforcement. This reduces expected crime and increases the likelihood of alcohol consumption. But over time, maintaining a weak law enforcement becomes costlier because overall crime will be rising fast as the alcohol market catches up and its associated crime networks develop over time (as captured by $k(T_t)$). This individual is constrained by a lack of independent policy instruments; by maintaining a low level of Prohibition enforcement, she is at the same time reducing overall crime enforcement. Moreover, the trade-off is more demanding the wetter the community, because a median voter in a wet community is more likely to have a liberal stand on Prohibition and be pessimistic about its effects, while facing a larger alcohol market.

On the other hand, the evolution of preferred law enforcement under Prohibition is determined by the difference between moral views and beliefs, both in their common ($z - \Omega_0 \theta_t^C$) and individual ($\theta_t^{med}$) components. From the martingale property of the stochastic updating process, $\Omega_t \theta_t^C \rightarrow a.s. \theta$. Nevertheless, the informativeness of signals, as measured by $\omega_t$, is increasing in law enforcement. Thus, early law enforcement choices are likely to be low, making early signals uninformative. Moreover, $\omega_t$ is also increasing in $k(T_t)$, so the relatively small alcohol market of early Prohibition years also reduces the informativeness of signals for a given level of law enforcement. As a result, learning should be slow during the first years under Prohibition, implying that the incentives to increase law enforcement as the alcohol market catches up are likely to dominate the incentives to reduce law enforcement due to changes in beliefs. As the market converges to its pre-Prohibition size (recall $k(T_t) \rightarrow 1$ as $T$ increases), and the increased levels of law enforcement increase the precision of the signals, learning will be faster and incentives to shrink law enforcement due to an increasing sequence of beliefs $\theta_t$ should dominate. After its initial fall, an invert U-shaped pattern of law enforcement should be observed.

Now, notice the presence of the term $k(T_t)$ in equation (18). Ceteris paribus, law enforcement should fall discretely right after Prohibition is introduced. This is the optimal response to the sharp contraction of the alcohol supply when it is closed on impact. Not only is the potential for crime small because the size of the alcohol market is smaller, but the marginal moral disutility of reducing law enforcement is also low because the alcohol market has sharply contracted, making it optimal to reduce policing. This increases the consumption of other public goods $G_t$ and the private utility of alcohol consumption for wet individuals. Moreover, if prior beliefs for the median individual are such that $\theta_0 > 0$, her ideal choice of law enforcement would fall even further because she is pessimistic about the response of crime to Prohibition.

In addition to these time-series predictions, equation (18) also makes predictions about the cross-sectional variation in law enforcement and learning. Specifically, variation in average moral views and alcohol market sizes across cities should interact with the evolution of beliefs. From equation (18), individuals in communities with higher average moral disutility (larger $z$) should be less sensitive
to changes in beliefs than individuals in communities where mean moral views are more favorable to alcohol. Thus, morally drier communities should be more reluctant to change law enforcement as learning takes place. This suggests a nuanced inverted U-shape pattern of ideal law enforcement in relatively dry cities. Of course, such a pattern across cities could be alternatively interpreted as arising from behavioral differences in the ability to learn, between individuals with differing moral views. This model can accommodate equilibrium differences in response to learning while still fully assuming rational individuals.

The correlation parameter $\rho$ has interesting implications in the model. A high correlation between individual moral views and prior biases implies that relative to no-Prohibition years, during Prohibition the decisive voter’s preferences will be more extremist, so that an amplification in the difference between the equilibrium choices of drier and wetter cities should be observed. Conversely, if this correlation is low, the average draw of $\theta_{med}$ will be very similar in Prohibition and no Prohibition periods, so that changes in law enforcement should not vary significantly between dry and wet communities when the legal standard is reformed.

5.6.2 Crime

Equations (10) and (11) are intended to capture some key features of the relationship between crime and the alcohol market. In the baseline equation for crime (10), I introduce $\Theta_S$, a scale parameter at the state level, to capture the large differences in the homicide rate levels across states. Following the claims of Prohibitionists, who argued that alcohol consumption was a source of criminality and social disruptions, I also allow it to vary with the size of the alcohol market.

Regarding Prohibition-related crime, a main reason why large cities were forced to maintain high police enforcement levels during Prohibition was their large potential for criminality, if policing were to be weakened. This points to a central conflict that arises in the context of Prohibition. The enforcement of a prohibitionist legal standard creates a non-separability between the objectives of enforcing Prohibition and of controlling crime. The instruments for the enforcement of Prohibition, mainly policing and judicial prosecution, are the same used to fight crime at the local level. If the enforcement of Prohibition creates crime, a community that does not favor Prohibition is unable to reduce law enforcement because it cannot be weakened without, at the same time, weakening overall crime enforcement. This is an especially binding constraint in relatively wet communities where criminality is more responsive to falls in crime enforcement, and motivates the functional form in equation (11).

It is frequently argued that crime increases during Prohibition were due to a shift of resources from crime protection to Prohibition enforcement. But this would predict exactly the opposite patterns to those observed in the data. It cannot explain why the steepest increases in crime and law enforcement were observed precisely in the wettest cities in the United States, since it would imply that relatively wet communities, strongly opposed to Prohibition, should have kept their Prohibition enforcement at very low levels and their crime enforcement resources high. This would have avoided a rise in criminality, and would have allowed the black market to operate with relative freedom. On the other hand, if the enforcement of Prohibition cannot be fully separated from overall crime enforcement, then wet communities must have been unable to reduce Prohibition enforcement.
Although likely to have a median voter more willing to invest in Prohibition enforcement, relatively dry communities faced smaller alcohol markets. Thus, they faced a lower potential for crime increases if law enforcement were to be weakened. These predictions are consistent with the patterns in the data\textsuperscript{32}.

Equation (11) also assumes that crime under Prohibition is a linear function of $\theta$. This is a relatively weak assumption, given that even if it is not linear, equation (11) could be seen as a first order approximation to any other nonlinear structural relationship between $q_t^p$ and the wedge in the alcohol market arising from Prohibition enforcement. Under such an interpretation, the error term would be capturing approximation error. Thus, any misspecification of this relationship should show up in the standard errors of the parameter estimates of equation (11).

5.6.3 Drunkenness Arrests

The functional form specifying the relationship between the drunkenness arrest rate and law enforcement implicitly assumes that throughout the relevant range of law enforcement intensities, the alcohol supply falls at a faster rate than that at which the arrest probability increases. It is adopted for simplicity only, since it makes the derivation of the conditional likelihood more straightforward, by allowing the mapping from unobserved variables to outcomes to be one-to-one for the whole range of outcomes. Moreover, the data suggests this is a reasonable assumption, since variation in law enforcement is only mildly correlated with variation in the drunkenness arrest rate, while the timing at which we know the market must have contracted is highly correlated with it across the sample.

5.6.4 Learning

In the model each dynasty gets a specific bias $\xi^i$, which is analogous to assuming heterogeneous priors in the population. Following Sethi and Yildiz (2009), $\xi^i$ can represent all the information which individual $i$ finds relevant about $\theta$, but is seen as irrelevant for everybody else. The historical literature has emphasized that initial public opinion regarding the effects of Prohibition was extremely optimistic. These biases came from two main sources - some relatively successful experiences of States that underwent Prohibition in the second half of the 19th century, and more importantly, the massive wave of prohibitionist campaigning and lobbying of the ASL and the WCTU in the decades prior to the adoption of nationwide Prohibition (See Asbury (1950); Blocker (1989); Foster (2002); Okrent (2010); Szymansky (2003)). I also assume that both wets and drys get their draw of $(\zeta^i, \xi^i)$ from same distribution. This is just a simplifying assumption since, for example, if wet individuals were to get their draw from a mean-shifted distribution, it would be isomorphic to increasing the difference in the marginal utility of private alcohol consumption between wet and dry voters.

\textsuperscript{32}This is not to say that a crowding-out of crime enforcement did not take place as communities increased the resources allocated to Prohibition enforcement. Indeed, the widely acknowledged congestion in judicial courts due to Prohibition-related cases is a good example of how it did to some extent shift resources away from overall crime enforcement. The argument here is just that the crowding-out had second order effects relative to the problem arising from the inherent difficulty in separating the enforcement of overall crime and of alcohol Prohibition.
Individuals only learn based on local information. This is in opposition to the experimentation literature where learning takes places from neighbors (for example, see Buera et al. (2010)). In the context of Prohibition it is likely that individuals were observing the crime outcomes of other cities. Nonetheless, it is very unlikely that they also observed the local law enforcement choices of other communities. Even if individuals believed that the effects of Prohibition were homogeneous across cities, a signal coming from a city from which law enforcement is not observed is void of informational content. Thus, in this model learning relies on local information exclusively, not only because it is likely that people recognized that Prohibition should have different effects in different communities, but also because learning from signals emerging from unknown law enforcement decisions is not possible without additional information. The endogenous nature of signals in this model justifies that learning should take place based exclusively on local information.

5.6.5 Political Environment

The political equilibrium of this model relies on two main assumptions. First, on simple majority voting as the collective choice mechanism. In the context of Prohibition in the United States, bipartisan political competition and a strong involvement of citizens in local politics were prevalent both at the local and federal levels. Indeed, political competition was much weaker in the South during the 1910-1930s, where the Democratic Party had a fairly generalized control of political power. Nevertheless, alcohol Prohibition as a political issue actually increased party competition by making dry voters, who were highly mobilized, involved in politics, and constantly motivated by dry organizations, pivotal.

Second, on the absence of any strategic experimentation considerations by voters. Because individuals live for only one period, they simply vote for the level of law enforcement which maximizes their current payoff given their present belief. In a more complex model, one could imagine long-lived or intergenerationally-altruistic voters making strategic voting decisions to induce experimentation in the collective choice of law enforcement level. In the context of Prohibition this is highly unlikely for several reasons. Foremost, local politicians’ incentives to experiment with law enforcement were very weak, since adverse criminality outcomes derived from “wrong choices” were likely to hurt their political careers. Indeed, as in any other experimentation setting, experimenting creates positive externalities since learning today benefits not only current, but also future constituencies, and thus, will in general be undersupplied by current constituencies (or politicians). Moreover, in voting environments, incentives to vote for experimentation (in the context of Prohibition higher levels of law enforcement) are weakened by the fact that pivotal voters under the present distribution of beliefs are likely to lose their decisive position after large changes in beliefs induced by experimentation (Strulovici (2010)).

33 A good example of how competition for the dry vote in the South did increase the competitiveness of local politics was the 1910 Tennessee gubernatorial election. The unwillingness of the incumbent Democratic governor Patterson to enforce the 1909 State Constitutional Amendment introducing Prohibition (after vetoing the Amendment and having his veto overridden by the legislature) alienated a dry fraction of the Democratic party, even after he stepped down for reelection. After more than 30 years in which the Republican party had not occupied Tennessee’s gubernatorial office, Republican candidate Ben Hooper won the election on a prohibitionist platform (See Isaac (1965) for a historically detailed account of Prohibition politics in Tennessee).
6 Structural Estimation

The equilibrium-political economy model developed in the previous section is characterized by three equilibrium relationships and the dynamic path of beliefs implied by Bayesian updating, which constitute the Data Generating Process (DGP) and can be directly used for estimation (recall that \(k(0) = 1\), and \(c\) indexes cities):

\[
q_{ct} = \Theta_S + k(\tau_c)\mu_c e^{\alpha_c p_{ct}} + P_{ct} \theta_c k(\tau_c) \mu_c [1 - e^{\alpha_c p_{ct}}] + \varepsilon_{ct}
\]

\[
d_{ct} = \frac{\mu_c k(\tau_c)}{\chi + e^{\alpha_c p_{ct}}}
\]

\[
p_{ct} = \frac{1}{\alpha_c - 1} \left\{ \ln [\alpha_c k(\tau_c)] + \ln \left[ \frac{\alpha_c}{\alpha_c + b} \left( z_{ct} - P_{ct} \Omega_c \theta_{ct}^C + 1 \right) + (1 - P_{ct}) \theta_{ct}^N + P_{ct} \theta_{ct}^P \right] - 1 \right\}
\]

\[
\theta_{ct}^C = \frac{1}{\sigma_{\xi}^2} B_c + \frac{1}{\sigma_{\eta}^2} \sum_{s=0}^{t-1} [q_{cs} - \Theta_S - \mu_c s k(\tau_c) e^{\alpha_c p_{cs}}] \omega_{cs}
\]

where \(\theta_{ct}^j\), \(j = N, P\) are distributed according to the densities derived in Appendix 2. In equations (19) and (20) the sources of randomness are \(\varepsilon_{ct}\) and \(\mu_c\) respectively; on the other hand, equilibrium police enforcement (equation (21)) was derived as a deterministic function. While mean morality in the community is part of each individual’s moral view, as an econometrician I can only estimate it. Thus, for estimation I will assume that \(z_{ct}\) is a normally distributed random variable with mean \(\bar{z}_{ct}\) and variance \(\sigma_z^2\): \(z_{ct} \sim N(\bar{z}_{ct}, \sigma_z^2)\). Although at the individual level moral views are fixed over time (in the model this is actually also true at the dynasty level), average moral views in the city will vary as the demographic/religious distribution of the population changes. This is particularly relevant during the early decades of the Twentieth century, when both European immigration to the U.S. and internal migration to the West and from the South to the North were very dynamic. Because I will estimate mean moral views using observable heterogeneity (mainly the distribution of religious ascriptions), the stochastic component of this variable can be thought of as capturing measurement error, or any other sources of variation in average moral tastes for alcohol, which do not vary at the individual level (recall that individual-level moral shocks are unobservable, and incorporated in \(\phi^i\)).

Given that the parameters of the model are identified only up to scale, I will normalize the variance of individual moral shocks \(\zeta^i\) to 1. Interpretation of all other parameters will thus be relative to \(\zeta^i\).

I am interested in obtaining estimates of the parameters of this model, which will also allow me to directly compute estimates of the common component of belief sequences \(\{\Omega_{ct} \theta_{ct}^C \}_{t=1}^T\) and of the shape of the distribution of the median voter’s unobserved preferred enforcement type \(\theta_{ct}^j\). Parameters to be estimated are listed below:
6.1 The Likelihood Function

I estimate the equilibrium political-economy model developed above through Conditional Maximum Likelihood (CMLE). Conditional on the decisive voter’s $\varrho^{med}$, this economy is characterized by a system of equilibrium structural equations for crime, drunkenness arrests and police enforcement, plus an equation that pins down the learning dynamics of the common component of beliefs.

Individuals, who are assumed to know the model and its parameters, learn about $\theta_c$ by observing the realizations of the outcome vector $y_{ct} = (p_{ct}, d_{ct}, q_{ct})$. The system in (19)-(21) has a particularly convenient “triangular” structure, which moreover, justifies the learning process implied by Bayesian learning and specified in equation (14). Once $p_{ct}$ is realized, conditional on $\varrho^{med}$ individuals face no uncertainty coming from equation (21) (recall that individuals observe $z_{ct}$). Then, after $d_{ct}$ is realized, the realization of $\mu_{ct}$ can be exactly backed-up from equation (20). As a result, in equation (19) the only remaining uncertainty about crime comes from $\varepsilon_{ct}$ and beliefs about $\theta_c$, which is consistent with the conditional distribution of $q_{ct}$ being normal, and hence, allowing the learning process to be as specified in section 5.

In Appendix 3 I derive the conditional likelihood function for the observed realization of the vector $y_{ct} = (p_{ct}, d_{ct}, q_{ct})$. It is given by

$$L_{ct}(y_{ct}; \Theta_S, \theta_c, B_c, a_c, b, \alpha_c, \chi, \varepsilon_{ct}, k, \lambda, \sigma^2, \sigma^2_z | \varrho^{med}(P_{ct}), P_{ct}, \tau_{ct}) =$$

$$\frac{[g_\mu(y_{ct})]^{\alpha_c-1}(1 - g_\mu(y_{ct}))^{b-1}}{\int x^{\alpha_c-1}(1 - x)^{b-1} dx} \exp \left( - \frac{1}{2\sigma^2_q} g_e(y_{ct})^2 \right) \exp \left( - \frac{1}{2\sigma^2_z} (g_z(y_{ct}; \varrho^{med}(P_{ct})) - \varepsilon_{ct})^2 \right) \frac{\partial g_\mu(y_{ct})}{\partial d} \frac{\partial g_z(y_{ct})}{\partial p}$$

(23)

where the expressions for $g_e(y_{ct})$, $g_\mu(y_{ct})$, and $g_z(y_{ct}; \varrho^{med}(P_{ct}))$ are given in the appendix. It is the product of a beta density coming from the distribution of the alcohol market size $\mu_t$, two normal distributions coming from the shocks to the crime rate and the random variation in mean moral views, and the relevant jacobian of the transformation. Central to identification, discussed further
below, the likelihood varies with $P_{ct}$. Prohibition introduces a structural change in the DGP, since a new nexus between law enforcement and criminality arises under Prohibition. A second key aspect of the model is that the DGP is dynamic; the vector of endogenous outcomes $y_{ct}$ depends upon previous values of itself. In this model, the dynamic component comes, of course, from learning. The equilibrium choice of law enforcement at time $t$, $p_{ct}$, is a function of the current updated beliefs about $\theta_c$, which depend on the whole sequence of previous realizations of the crime rate during Prohibition years $\{q_{ct}\}_{s=1}^{T}$. In the likelihood (equation (23)), the dynamic component enters through $g_c(y_{ct}; g^med(P_{ct}))$ exclusively.

While $\sigma^2_\theta, \sigma^2_\varepsilon, \chi, \kappa,$ and $\lambda$ are assumed constant across cities, I allow the parameters in the likelihood function above to vary with observable community characteristics as follows:

- $\Theta_S = x_S^B \Sigma$, where $x_S^B$ includes state-level dummies.
- $\theta_c = x_c^B \Lambda$, where $x_c^B$ includes border cities, South, state-capitals indicators, average demographics, and a constant.
- $B_c = x_c^B \Xi$, where $x_c^B$ is a vector containing the initial religious ascriptions distribution and a constant.
- $a_c = x_c^B \Gamma_a$ and $b$, where $x_c^B$ includes average demographics, average religious ascriptions, average population, and a constant, and $b$ is constant across cities.
- $\alpha_{ct} = x_{ct}^M \Psi$, where $x_{ct}^M$ is a vector of legal enforcement variables (and a constant) such as the number of state-level dry laws in place (in years when the city is not under Prohibition these are the only source of restrictions on the alcohol market), a dummy equal to one when a city’s state has a Prohibition enforcement law (during Prohibition), and other variables which might be correlated with federal law enforcement (a border city dummy, a Bureau of Prohibition period dummy, and dummies for the different Prohibition districts).
- $\varepsilon_{ct} = x_{ct}^M \Pi$, where $x_{ct}^M$ is a vector of containing the religious ascriptions distribution, and a constant.

Let $\beta = (\Sigma, \Lambda, \Xi, \Gamma_a, b, \chi, \kappa, \lambda, \Psi, \Pi, \sigma^2_\theta, \sigma^2_\varepsilon)$, and $x_{ct} \equiv (x_S^B, x_c^B, x_{ct}^B, x_{ct}^M, x_c^B, x_{ct}^B)$. The conditional likelihood can be more compactly written as $L_c(y_{ct}; y_{ct-1}, x_{ct}, \beta | g_\text{med}, P_{ct}, \tau_{ct})$, which makes its dynamic nature explicit. Once the dynamic process is correctly specified (in this case the Bayesian learning assumption) and incorporated into the likelihood function, the density of the outcome vector $y_{ct}$ only depends on $y_{ct-1}$ through the learning channel, and hence the DGP is dynamically complete (See Wooldridge (2002, p. 412)). As a result, conditional on $y_{ct-1}$, the $y_{ct}$ are independently distributed. Thus, the conditional likelihood for a given observation $y_{ct} = (y_{ct1}, y_{ct2}, \ldots, y_{ctT})'$ is given by $L_c(y_{ct}; \beta | g^\text{med}, P_{ct}, \tau_{ct}) = \prod_T L_c(y_{ct}; y_{ct-1}, x_{ct}, \beta | g^\text{med}(P_{ct}), P_{ct}, \tau_{ct})$, where $g^\text{med}$ is drawn

---

34While the first moment of the beta distribution is determined by the difference between $a$ and $b$, its second moment is symmetrically decreasing in the magnitude of both $a$ and $b$. Thus, allowing one of the parameters to depend on demographics and the religious distribution, while making the other one common across cities, allows this source of variation to identify the first and second moments. Allowing $b$ to vary across cities could only increase the fit of the model. (This follows Coate and Conlin (2004)). Because I am assuming that $a$ and $b$ are constant across time for each city, I use the time-averaged values of the demographic and religious variables.
from $f_{\theta_{med}}(\theta_{med}; a_c, b, \sigma_2^2, \sigma_3^2, \rho)$ during Prohibition years, and from $f_{\theta_{med}}(\theta_{med}; a_c, b, \sigma_2^2)$ during years without Prohibition. Given that the $\theta_{j}^{med}$ are unobserved, it is necessary to integrate them out from the conditional likelihood, using their derived equilibrium densities. Estimates of $(\beta, \sigma_{2}^2, \rho)$ are obtained from the following program:

$$\max_{\beta, \sigma_{2}^2, \rho} \sum_c \ln \left\{ \prod_{t=1}^{T} L_{ct}(y_{ct}; y_{ct-1}, x_{ct}, \beta|\theta_{med}(P_{ct})) \right\} f_{\theta_{med}}(\theta_{med}; a_c, b, \sigma_2^2, \rho, P_{ct})d\theta_{med}$$  \hspace{1cm} (24)

As a final observation, dynamic models estimated by MLE usually face an “initial conditions” problem, arising from the fact that the observation for the first year in the sample depends upon an unobserved realization of the endogenous variable (See Wooldridge (2005)). In this model such a problem does not arise because for years under no Prohibition, the likelihood function does not depend on previous realizations of $y_c$, and for the first period under Prohibition, the learning model implies that beliefs are exclusively based on the prior $\theta_{ct}$, which is not a function of $y_{ct-1}$ either. For all subsequent years under Prohibition, the relevant lagged information is available. Of course, this relies on having a sample covering for every observation, at least one year under no Prohibition.

Ideally, estimation of the model would cover the whole period; unfortunately, the drunkenness arrests data is only available for the years 1911-1929. Because this variable is necessary in the estimation to identify the alcohol market dynamics, I estimate the structural model for that period. Nevertheless, this imposes some discipline since it allows performing an out of sample exercise with the model’s estimates to predict the observed data for the period 1930-1936. Thus, the sample used for the structural estimation consists of a fully balanced panel of 66 cities from 31 different U.S. states, for the nineteen year period 1911-1929. This makes a total of 1,254 city-cross-year observations for which the homicide rate, the drunkenness arrest rate, the police expenditure share, and all of the demographic, religious and legal enforcement variables are available.

The only endogenous variable with a strong trend throughout the sample period, unaccounted for in the model, is the police expenditure share. Closer examination of the raw data reveals that this downwards trend is the result of a strongly increasing trend in total public spending across all cities in the United States during those years. Thus, for estimation I use the de-trended police expenditure share as the measure for $p_{ct}^{'}$. As the crime outcome measure, I use the natural logarithm of the homicide rate, which standardizes the variance in homicide rates across cities, and is consistent with the shocks in equation (19) being normally distributed, and drawn from the same distribution across cities. Table A5.1 presents the list of cities included in the estimation and discusses the data further.

$^{35}$While the average annual growth rate of total public spending in the sample was 5.6% (s.e. = 2.2%), the same number for police expenditure was only 3.7% (s.e.=2.5%). To obtain the detrended police share variable I ran a regression of the raw police expenditure share $p_{ct}^{'}$ for each city in the sample, on a city-specific linear time-trend and city effects, and no constant: $p_{ct}^{'} = \alpha_c + \beta_{ct} + \epsilon_{ct}$. I then compute the detrended police share as $p_{ct} = \alpha_c + \bar{\epsilon}_{ct}$. Of course, this is equivalent to running a separate regression for each city.
6.2 Moments Identifying the Parameters in the Model

In this subsection I briefly discuss the relevant moments identifying the different parameters of the model. The structural elasticity of crime to the adoption of Prohibition, $\theta_c$, is a function of city characteristics. Thus it is identified off the covariation in the homicide rate between cities with similar characteristics, and from the time-series variation in the homicide rate between periods under no Prohibition and periods under Prohibition. As previously noted, functional form is not key for the identification of $\theta_c$, given that equation (19) can always be taken as a first order linear approximation to any monotonic relationship between the homicide rate and Prohibition enforcement.

Parameters $a_c$ and $b$ are identified off the residual variation in drunkenness arrests, once law enforcement and the catch-up of the alcohol supply have been accounted for. Since variation in law enforcement is correlated with variation in the availability of alcohol, the “wet” share cannot be identified from the drunkenness arrests data without additional information. This additional information comes from two sources: the variation in the homicide rate, by exploiting the fact that in a given city the drunkenness arrests and the homicide rate jointly covary with law enforcement, and the dynamics of the supply of alcohol under Prohibition, which the model assumes takes a particular functional form and is common across cities. It relies on two assumptions. First, that the baseline arrest probability, determined by $\chi$, is constant over time, so that any changes in arrests between no-Prohibition and Prohibition years come solely from changes in law enforcement intensity, and not, for example, from changes in the “arrest technology”. Second, that preferences over private alcohol consumption are independent of Prohibition status. Although a strong assumption in the context of Prohibition, a priori it is unclear in which direction tastes for alcohol might change when the community is under Prohibition. On the one hand, citizens might derive utility from abiding by the law, no matter what restrictions it imposes on their individual freedoms; on the other hand, they also could be subject to a “forbidden fruit” effect, where utility derived from a prohibited activity increases precisely because it is forbidden. Relatedly, since the baseline drunkenness arrest probability $\chi$ is assumed constant over time and across cities, $\chi$ is identified from the variation in arrest rates that is common across cities over time.

Regarding $\alpha_{ct}$, the model assumes that the dynamics of the legal standard are exogenous from the point of view of the city. Although citizens were voting both for local law enforcement and for the state and federal legal standards, the assumption is that within a state or the Country as a whole, each city is too small to affect the equilibrium choice of legislation. This seems like a natural assumption, given that citizens in more rural areas were more strongly in favor of Prohibition. Indeed, many urban citizens of the United States saw the introduction of Prohibition as an intrusion from rural interests. Even in a state like New York, the pressure from Upstate voters set restrictions on the ability of New York City to dismantle Prohibition completely. At some level, this paper is about the effects of the imposition of a legal standard over communities where a large fraction of their members were in opposition to it. Thus, identification of $\alpha_{ct}$ comes from the common variation in drunkenness arrests and the homicide rate induced by changes in state-level legislation.

Identification of the city-specific collective prior, $B_c$, comes from early years under Prohibition, when the community choice of police enforcement closely follows prior beliefs. The larger the initial biases, the larger the gap between the observed police enforcement choice and what the optimal choice would
be under perfect information. Because the model estimates $\theta$, it implicitly provides a measure of how “off” law enforcement decisions were during early Prohibition years. In the model, the correlation of prior beliefs across cities depends on the distribution of religious ascriptions. Thus, the covariation between the gap from “optimal” law enforcement and the distribution of initial religious ascriptions identifies $B_c$.

On the other hand, the $\kappa$ and $\lambda$ parameters are identified from the common time-series residual variation in drunkenness arrests across cities, unaccounted for by changes in law enforcement or by changes in the wet share. The identification of these parameters relies strongly on the functional form I assume for the alcohol supply “catch-up” process, and the assumption that this catch-up is common for all cities in the sample. Nevertheless, the functional form in equation (8) is very flexible and can accommodate a wide variety of nonlinear trends.

Average moral views $\bar{z}$, which are function of the religious ascription distribution in the community are identified, from equation (21), from the variation in the police expenditure share which is uncorrelated with changes in beliefs, the alcohol market size, or dry legislation. Because the alcohol supply and beliefs change over time only during Prohibition years, the identification of $\bar{z}$ comes from the variation in law enforcement which is common for the city before and during Prohibition. On the other hand $\sigma^2_z$, the second moment of the distribution of $z_{ct}$, is identified directly from the sample variation in police enforcement that is common across cities.

Finally, $\sigma^2_x$ and $\rho$ are identified in the model from the change in the shape of the estimated density of $\theta_{med}$ between no-Prohibition and Prohibition years, as figure 12 illustrates. As the variance of the distribution of biases decreases, the density of $\theta_{med}^p$ shifts to the left relative to the density of $\theta_{med}^N$. This is because the weight on the prior is larger, and as a result, law enforcement choices give more weight, on average, to individuals’ biases. The effect on the density of $\theta_{med}^p$ is similar as $\rho$ increases in magnitude because a larger $\rho$ (in absolute value) magnifies the differential law enforcement decision of dry cities relative to wet communities, increasing the variance of the distribution of $\theta_{med}^p$ relative to the distribution of $\theta_{med}^N$. The reason is that if moral views $\zeta^i$ and belief biases $\xi^i$ are correlated, this should have no effect on the preferences of the median voter when the city is not under Prohibition. During Prohibition, beliefs do shift the preferred police expenditure relative to no Prohibition periods, and the larger the correlation is (in absolute value), the larger the difference in the choice of optimal law enforcement between individuals with differing moral views. As $\rho$ increases in magnitude, the density under Prohibition shifts mass to the left, making lower values of police expenditure more likely. Thus, $\rho$ is of special interest in the estimation since it is identified off the channel stressed the most in Section 4 (the differential law enforcement choices between communities with varying moral views), highlighting the importance of the unobserved sources of heterogeneity in preferences over law enforcement for their dynamics during Prohibition.

6.3 Fit and Results

This section presents the estimation results from the CMLE. I start discussing the overall fit of the model’s benchmark specification, and subsequently discuss the parameter estimates. To provide a general idea of the fit of the model across cities, panel A in figure 8 presents the city-level scatterplots of the average (over time) observed and predicted outcome variables, together with the
45 degree lines over which a perfect fit would obtain. The predictions are computed directly from equations (19)-(22), where I use the estimated expected value for the wet share \( \mu_{ct} \) for each city, \( a_c / (a_c + b) \), in the computation of the belief sequences, the predicted drunkenness arrest rate, and the predicted log homicide rate. For the predicted police shares, I use the mean value of the \( g_{med} \), which I calculate by integrating over the estimated equilibrium densities \( f_{\varphi_P} (\varphi_P; a_c, b, \sigma_{\varphi_P}) \) and \( f_{\varphi_N} (\varphi_N; a_c, b, \sigma_{\varphi_N}) \). The figures illustrate that the model does a fairly good job in fitting the variation across cities in the sample, especially for the drunkenness arrest rate and the homicide rate. The model tends to over-predict the small observed values and to under-predict large ones. The equation for the police share is no doubt the harder to fit, because preference heterogeneity, changes in the alcohol market size, and changes in beliefs are all interacting.

Regarding the time-series dimension, panel B in figure 8 presents the average (across cities) observed and predicted outcomes, for the sample years. For the three outcomes, the model is able to capture the joint evolution quite accurately, albeit with some differences in magnitudes. For example, it predicts a more pronounced fall in the police share than the one observed around the years 1920-1923, when the majority of cities were experiencing their first years under Prohibition. The apparent reason is that in the model, policing choices are quite sensitive to the size of the alcohol supply, and the impact effect of beliefs when cities enter into Prohibition is not large enough to counter the estimated fall in the alcohol supply. For the later years, the average predicted police share is around 0.1 percentage points larger than the observed. On the other hand, the predicted magnitude of the fall in the drunkenness arrest rate falls short from the one observed in the data between 1916 and 1920. The model is attributing a fraction of the fall in the drunkenness arrest rate to sampling variation from the distribution of the wet share \( \mu_{ct} \). The model also predicts the fall to begin somewhat later, around 1918. Finally, the last figure depicts the predicted log homicide rate, showing that the model overpredicts the level of the homicide rate during the 1910s, and also predicts a smoother increase in this variable, compared to the rapid rise in homicides observed in the sample around 1920-1924. The reason for the overprediction of crime in early years is that I allow the alcohol market to have an effect on crime during the period without Prohibition. This suggests little or no room for an effect of the alcohol market on the homicide rate when Prohibition is not in place.

In addition, a way to assess the fit of the model is to look at the variability in the average moral views required to match the data. From equation (21), if the evolution of law enforcement, the alcohol supply, beliefs, and the change in the distribution of \( g_{med}^i \) are able to match the police data closely, variation in average moral views \( z_{ct} \) over time should be small. In the model, the estimated \( \sigma_z^2 = 0.06 \) (s.e = 0.03), which relative to \( \sigma_z^2 \), the variance of individual moral tastes \( \zeta^i \) (normalized to 1), is quite small. Overall, the estimates suggest that the mechanisms highlighted in the model...

---

36 The reason why the model predicts larger drunkenness arrests in the years in which these fall sharply is that by making the fall in the market supply larger, the fit of the police enforcement equation would be reduced because in the data, policing is not as sensitive to the dynamics of the alcohol supply. The large variability in drunkenness arrests could also be captured with a larger variance in the “wet share” distribution \( \mu \). Nevertheless, because the distribution of the data is positively skewed, an increase in the variance would require a larger estimate of \( a_c \), which would imply a higher elasticity of the equilibrium police share to moral views and beliefs (see equation 21), reducing the ability of the model to fit the observed police outcomes.
capture a significant fraction of the joint variation in the data, despite the relatively small sample size.

6.3.1 Estimates

Estimates of the covariates from the model are presented in table 4, and table 5 presents the implied average estimated values of the main parameters of the model, based on the coefficient estimates. Standard errors for the coefficients are computed through a bootstrap. Among the covariates for $a_c$, the demographic variables all have positive and significant coefficients as expected; cities with larger populations 15-44 years of age, larger foreign white populations, and larger black populations, have larger “wet shares”. On the other hand, most of the coefficients on the religious ascriptions are unprecisely estimated, although the estimates for the religions traditionally considered as “wet” (Orthodox, Lutheran, and Catholic) are significant and positive. Finally, population size does not explain variation in the wet share. Together, the average estimate of $a_c$ across cities is 3.16 and is 8.66 for $b$, implying that the average “wet share” is around 0.267. Since $a_c$ varies little across cities (its standard deviation is 0.055), the model predicts very similar sizes of the “drinking population” across cities.

Looking at the covariates for average moral views $\bar{z}_c$, Baptist, Evangelical and Methodist shares do significantly increase average moral views. Surprisingly, the coefficient for the Catholic share is large in magnitude (0.81), but imprecisely estimated. Looking at the covariates for $\alpha_{ct}$, the alcohol-related laws variable is insignificant (point estimate = 0.11, s.e. = 0.25), suggesting that changes in dry legislation had little effect in making policing more effective for Prohibition enforcement. On the other hand, the coefficient on the Enforcement Law dummy is negative and quite significant (point estimate = -0.89, s.e. = 0.21), suggesting that the repeal of state-level Prohibition enforcement laws made policing more effective for crime enforcement. This might be driven by the unwillingness of local authorities to enforce Prohibition laws which they oppose. Regarding the Prohibition Unit indicators, out of which the New York Unit is omitted, all other Units except for the San Francisco and Los Angeles ones have negative estimated coefficients. This is consistent with historical observation that federal law enforcement was especially focused around the mid-Atlantic “wet” states, and with a relatively dry coastal California, which likely made a given amount of policing more effective in reducing the alcohol market.

Of central interest are the model’s estimates of $\theta_c$, the structural “elasticity” of Prohibition enforcement to crime. The average $\theta_c$ is 1.37, and figure 9 presents the distribution of estimated $\theta_c$’s across cities. These range from around 0.8 to 1.6. At the means of the police share $p_{ct}$ and the estimated parameters, it implies that the average city saw an increase in the homicide rate of around 23% during Prohibition. Among the estimates for its covariates, the estimate for the border indicator (Canadian, Mexican or coastal city) is negative and significant (point estimate = -0.302, s.e. = 0.101). Given the accounts of huge amount of smuggling during the Prohibition years, this is at first puzzling, but actually consistent with my discussion above, about borders having the effect

\[ 0.23 = \exp(1.37 \times 0.8 \times 0.267 \times \{1 - \exp(-3.43 \times 0.36)\}) - 1. \]

\[ \theta_c = 1.37, \quad \text{and} \quad p_{ct} = 0.36. \]
of increasing the availability of alcohol for a given level of law enforcement, and thus, reducing the incentives for Prohibition-related crime to arise.

Estimates for the covariates of the Prior $B_c$ are also presented in table 4. Baptist and Methodist shares have the largest (in magnitude) estimated significant coefficients, implying that cities with larger fractions of members of these religions initially did have more optimistic Priors about the effects Prohibition would bring. The Catholic share also has a coefficient of large magnitude, but once again its standard error is quite large. On the other hand, the estimates for the second moments of the joint distribution of individual biases and moral views (see equation (6)) also are of interest. The variance of biases $\sigma_\xi^2$ is estimated to be 0.34, implying that the variation in individual moral views (recall its variance $\sigma_\xi^2$ was normalized to 1) was significantly larger than variation in biases. In the model, the magnitude of $\sigma_\xi^2$ is a measure of how much weight people put on their prior beliefs, so that smaller values of $\sigma_\xi^2$ directly imply slower learning. Finally, $\rho$, the estimated correlation between prior biases ($\xi^i$) and moral views ($\zeta^j$) is $-0.49$ (s.e. = 0.09), suggesting that cities with constituencies more favorable to Prohibition did have much more optimistic beliefs about its effects.

An alternative way to see the correlation between moral views and beliefs from the model’s estimates is with a scatterplot of the estimated values of priors $B_c$ and average moral views $\bar{\zeta}$, for the cities in the sample. Figure 10 presents such a scatterplot, together with a simple regression line. Its slope is $-0.25$ with a t-statistic of $-6.36$. Thus, even in this sample of relatively large cities, average prior beliefs and moral views were negatively correlated. In particular, the model predicts negative values of prior beliefs for all cities in the sample. This is because the cities observed a relative increase in policing in the early years under Prohibition (see figure 8), which in the model is driven by optimistic priors. Because early on during Prohibition learning is slow, the sharp fall in the alcohol supply more than offsets the average increase in beliefs, explaining the subsequent fall in policing observed in the data.

The parameter estimates from table 5 also allow a quantitative characterization of the structural relationships specified in the model. In particular, the estimates for $\kappa$ and $\lambda$ from equation (8) ($0.26$ and $0.25$) imply that at its lowest point, the supply of alcohol was on average 68% its pre-Prohibition level, and that this minimum was attained around 3.75 years after the introduction of Prohibition. Together with this estimated function for the alcohol supply catch up, figure 11 presents the estimated drunkenness arrest conditional probability, and the estimated percent increase in the homicide rate due to Prohibition, both as a function of police expenditure. The three graphs in the figure present an illustrative picture of the costs and benefits of Prohibition. Prohibition was able to shrink the alcohol supply by about 35%, but only for a relatively short period of time. While increasing policing would increase arrests for drunkenness, the slope is not very steep. Considering that the average standard deviation of (normalized) police shares in the sample is 0.044, a whole standard deviation increase in the police share would at most increase the arrest probability by 3%. In sharp contrast, the same increase in policing during Prohibition would imply that the homicide rate would move from being 23 to 24.6% higher under Prohibition.\footnote{The minimum of equation 8 is attained at $\tau = 1/\kappa$.}

\footnote{Thus, using the average parameter estimates from table 5, the estimated arrest probability is computed as $Pr(Arrest(p) = \frac{e^{3.43p}}{1 + e^{3.43p}})$, and the estimated proportional increase in the homicide rate under Prohibition is computed as $\Delta Q(p) = e^{1.37 + 0.8 \times 0.267 \times (1 - e^{-3.43p})}$, for $k(\tau) = 0.8$.}

\footnote{Average (normalized) police share is around $p = 0.36$. Thus, the increase in the arrest probability induced}
Finally, the estimated shapes of the distributions of the unobserved $\theta^\text{med}_t$ (the median voter’s “type” under no Prohibition and under Prohibition) can be directly derived from the parameter estimates of the structural model, by plugging the estimates of $a_c$, $b$, $\sigma^2_\zeta$, $\sigma^2_\xi$, and $\rho$ in equation (32) from Appendix 2. Figure 12 plots both densities, for the mean values of the parameter estimates, and for the first year under Prohibition (when $\Omega_t = \sigma^2_\xi$). The difference in skewness between the distribution under Prohibition and under no Prohibition is what identifies $\rho$ in the model. This is because the larger (in magnitude) the correlation between moral views and belief biases, the larger the average difference in policing choices that a median voter would make, when passing from no Prohibition to Prohibition. Also, as $t$ increases, $\Omega_t \to 0$, so that the density under Prohibition converges to the density under no Prohibition.

6.3.2 Learning

In this subsection I discuss the estimation results related to learning. Recall the model estimates a relatively low variance of individual belief biases $\sigma^2_\xi$. This is the main exogenous parameter affecting the speed of learning in the model, and is common across cities. Thus, differences across cities in the estimated speeds of learning are due directly to the variation in enforcement choices over time, which under normal updating, affect the informativeness of the signal. The reason for the relatively low estimate of $\sigma^2_\xi$ is that the model is estimated over the years 1911-1929, thus, excluding precisely the later years under Prohibition (1930-1936), in which the largest adjustments in law enforcement occurred. Nevertheless, there is substantial learning over the nineteen year period. Figure 13 graphs the evolution of the estimated empirical distribution of the common component of beliefs $\{\{\Omega_t \tilde{\theta}^\rho_{ct}\}_{c=1}^N\}_{t=1911}^{1929}$, derived directly from applying equation (22) iteratively using the estimated coefficients and the observed sequences of outcome variables. The outermost curves represent the 10th and 90th percentiles, the curves in between represent the 25th and 75th percentiles, and the middle curve represents the median of the estimated distribution. Of course, beliefs remain at the prior until cities fall under Prohibition status. In several cities, for some of the early Prohibition years, beliefs about $\theta_c$ actually fall slightly. After around 1923 though, the belief sequences are monotonically increasing for all cities, but there is substantial variation in the speed of belief updating. The figure also shows that despite the generalized increasing pessimism over the effects of Prohibition, the dispersion of beliefs actually increases over time. Mean common beliefs increase from the average prior $B_c = -1.31$ to a mean posterior of $-0.57$ in 1929, whereas the posterior median is only around $-0.75$. While the standard deviation of priors is 0.26, it is 0.68 for the 1929 posteriors. At some level, this is a natural implication of the model, given that each city is learning from its own experience exclusively, and that different cities had different structural values of $\theta_c$.

A key question is whether the differential evolution of beliefs across cities, is correlated with differences in their moral profiles. The reduced-form estimates already suggested that this is the case. Recall from Section 4 that during the first years under Prohibition, wetter cities had differentially lower levels of police enforcement. I argued there that this could be driven by the willingness of from increasing policing to 0.4 = 0.36 + 0.04 would be $\frac{\exp(3.43 \times 0.4)}{\exp(3.43 \times 0.4) + 1} - \frac{\exp(3.43 \times 0.36)}{\exp(3.43 \times 0.36) + 1} = 0.028$. On the other hand, the shift in the homicide rate goes from $\exp(1.37 	imes 0.8 \times 0.267 \times [1 - \exp(-3.43 \times 0.36)]) - 1 = 0.232$ to $\exp(1.37 \times 0.8 \times 0.267 \times [1 - \exp(-3.43 \times 0.4)]) - 1 = 0.246$ under Prohibition, by increasing policing by one standard deviation around its average.
more optimistic “dry” cities to invest in law enforcement. In the context of a learning model, dry cities should learn faster early on, given that their signals are more precise. The estimates here are consistent with that view; running a regression of the estimated 1929 posteriors on the estimated average moral views, and controlling for the estimated priors, the coefficient estimate on moral views is positive and has a t-statistic of 2.26. Thus, although the standard deviation of beliefs across cities increased over time, the incentives for differentially higher initial law enforcement in drier cities limited the extent of divergence in beliefs. This is also consistent with the fact that among the subset of relatively “wetter” communities, dry ones saw larger shifts of public opinion against Prohibition (see figure 7). Overall, the structural estimates of the evolution of beliefs are consistent with the correlations from the reduced-form analysis.

At the heart of the model is the endogenous evolution of outcomes due to rational learning. Thus, I end this subsection by estimating the model closing the learning channel, to assess the relative performance of a model where no learning occurs compared to the benchmark specification (this follows Buera et al. (2010)). Formally, this is equivalent to imposing the restriction $\sigma^2 = 0$, so that individuals never update their priors. A simple Likelihood Ratio test can be performed comparing the restricted No-Learning model with the benchmark model. The log-likelihood for the model without learning is 5,978.99, while the log-likelihood for the benchmark model is 6,560.77. Under the null hypothesis that the restricted and unrestricted models are indistinguishable,

$$LR = 2[logL(Benchmark) - logL(NoLearning)] \sim \chi^2_{701}$$

Assuming $\sigma^2 = 0$ implies a restriction in the police equation for each city, in every year under Prohibition except the first. There are 767 such observations, so the appropriate number of degrees of freedom for the test’s $\chi^2$ distribution is 701. While $LR = 1,163.55$, the 99% critical value is 791.03. Thus, the null can be rejected at any significance level.

6.4 Counterfactuals

To conclude, I exploit the model’s estimates to perform a series of counterfactual exercises. These should allow a further assessment of the model’s fit, and also provide general-equilibrium answers to questions of interest, which would be impossible to make in a partial equilibrium or reduced-form framework. First, I perform an “out of sample” prediction of the outcome variables for the years 1930-1936, using the parameter estimates. I then ask the following questions to the model: what would the evolution of outcomes have been under Prohibition, if average prior beliefs had been unbiased? How would Prohibition outcomes have evolved if society had been more radicalized? More polarized? Finally, I assess the implications of alternative political environments.

For the 66 cities in the sample, I run the regression $\Omega_{c,1929}\hat{\beta}_{1929} = \beta_0 + \beta_1 \bar{q} + \beta_2 \Omega_{c,1911}\hat{\theta}_{c,1911} + \varepsilon$. The estimated $\beta_1$ is 2.17 with a standard error of 0.96. I include the prior as a regressor to control for the fact that morally drier cities had more negative priors.
6.4.1 Out of Sample Prediction

Because of the unavailability of drunkenness arrests data for years after 1929, I am unable to estimate the model for the later Prohibition years. Thus, I make an out of sample prediction for the police and homicide outcomes during the 1930-1936 years, by using the MLE estimates on equations (19)-(22). This exercise is particularly meaningful because I do observe the police and homicide rate outcomes in that period, so I directly can assess the extent to which the model is able to capture the subsequent evolution of Prohibition during its final phase, and the first few years after its repeal. For this purpose, I take the estimated 1929 posterior beliefs for each city, and use them as the 1930 priors. I then compute iteratively the predicted equilibrium values of $p_{ct}$ from equation (21), and with this predicted police enforcement value, I then predict $q_{ct}$ from equation (19). To compute year $t$'s posterior from equation (22), I add a random shock drawn from a mean-zero normal distribution with variance equal to 0.277 (the MLE estimate for the variance of $\varepsilon$, $\sigma^2_\varepsilon$) to the predicted value of $q_{ct}$ and iteratively use this posterior to calculate year $t + 1$'s police choice and homicide rate. Constitutional Prohibition was repealed in the end of 1933, so belief updating actually stops after this year. Figure 14 presents graphs analogous to those in figure 8, comparing the “out of sample” average predicted values from the structural model, both in the time and city dimensions. Panel A shows the scatterplots of the 1930-1936 averages for each city. The horizontal axis has the observed values, while the vertical axis has the predicted values. The predictions for the homicide rate are again fairly close to the observed. For the policing data, the slope is significantly positive, but as in the predictions for the 1911-1929 period, the model is not able to capture all of the variability across cities. Looking at panel B, on the other hand, it captures the trend of both variables over time remarkably well, in particular the fall in both policing and the homicide rate during the last years of Prohibition, and the leveling off of both variables after repeal.

6.4.2 Changes in Prior Beliefs

The adoption of Prohibition in the U.S. would not have been possible based exclusively on moral motivations, since radically dry sectors did not constitute a large enough majority of the population. Its adoption required a large fraction of morally-indifferent voters to have optimistic beliefs about the effects of the policy. Thus, a natural question arises: what was the cost of these biased prior beliefs? The model can provide an answer to this question, by making the counterfactual exercise of assuming that priors were unbiased. Specifically, I assume that prior common beliefs in 1911 were unbiased, this is, that $B_c = \theta_c$. Using the estimated coefficients, I can then compute the predicted evolution of outcomes over time, and compare them to the model's predicted outcomes under the estimated biased priors.

The simulation results reveal several patterns. As expected, beliefs endogenously remain fairly unchanged over time, since the realized homicide outcomes are always close to the expected ones given the law enforcement choices. Police enforcement decisions, on the other hand, behave quite differently. In particular, cities would have avoided the early-Prohibition increases in policing, since in the absence of optimism about Prohibition’s effects, there are no incentives to increase law enforcement. Subsequently, policing decisions would have fallen sharply relative to the benchmark case, following the early contraction of the alcohol supply, and would have bounced back at a
relatively faster pace. In contrast, when beliefs are biased, learning makes this effect nuanced as the median voter finds it less attractive over time to maintain high levels of police enforcement. The model predicts that the median city would have reduced law enforcement to almost half the predicted law enforcement levels under biased beliefs. Thus, cities would have been much more radical in offsetting Prohibition with their local law enforcement choices. Variation across cities in law enforcement would have increased, on the other hand, because the variance in the distribution of Prohibition-related crime potential $\theta_c$ is larger than the variation in estimated priors. In addition, the model also suggests that the differences in the homicide rate relative to the biased-beliefs case would have been insignificant. This is because the inability to reduce Prohibition enforcement without concomitantly reducing overall crime enforcement implies that the relatively large fall in policing would allow for an increase in non-Prohibition related crime. Somewhat counterintuitively, this suggests that conditional on Prohibition been imposed, more accurate initial beliefs about its effects could have allowed the policy to remain in place longer, because large cities would have faced relatively similar crime outcomes, but lower police enforcement expenditures. Since beliefs would not have changed significantly, public opinion change would have been limited.

6.4.3 Radicalization and Polarization

The model also can address questions related to the distribution of preferences in society. I perform two simple exercises. I start by asking what the evolution of outcomes under Prohibition would have been, relative to the estimated benchmark model, under more radical moral views against alcohol consumption. This implies a higher degree of alignment between the prohibitive legal standard and preferences over its enforcement. Consequently, I increase each city’s estimated average moral view $\bar{z}_{et}$ by one or two standard deviations (the estimated $\sigma^2 = 0.06$), and compute the predicted sequences of outcomes under these changes. The model predicts that these radicalized communities would choose a constantly higher level of police enforcement (around 20% more for the one standard deviation increase, and around 36% more for the two standard deviation increase), but variation in police choices across cities would also be larger. Common beliefs would consequently be updated faster relative to the benchmark model’s predictions. Nevertheless, in this case it is unclear whether public opinion would turn against Prohibition as fast as it in fact did, given that across cities the decisive voter is more willing to restrict the alcohol market for a given belief profile.

Another exercise of interest is to increase the degree of polarization in society. By polarization here I mean increasing the average willingness to enforce prohibition, by raising $\bar{z}_{et}$, and at the same time increasing the demand for alcohol, by raising the mean of the distribution of $\mu$. Thus, just as in the counterfactual exercise above, I allow $\bar{z}_{et}$, but also $E[\mu]$, to increase by one or two standard deviations. The estimated standard deviation of the “wet share” $\mu$ is 0.123, while its mean is 0.267, so that a one standard deviation increase in the mean implies $E[\mu] = 0.38$. Holding $a$ fixed, such a shift in the distribution of the wet share can be achieved by reducing the value of $b$ to 5. For a two standard deviation increase in the mean of $\mu$, which implies $E[\mu] = 0.5$, a value of $b = 3.1$ achieves the same objective. The model predicts that the speed of learning during Prohibition increases very fast on the degree of polarization in society. The benchmark model’s estimated 1929 posterior common beliefs for the median city would have been reached by 1923 if both average
moral views and the average wet share were one standard deviation larger, and by 1921 if they were
two standard deviations larger. This outcome is the result of increased police enforcement levels
as the degree of polarization increases. Given the model’s parameter estimates, this occurs for two
reasons. First, more radical moral views increase the ideal choice of Prohibition enforcement across
the population. Moreover, because prior beliefs were initially relatively optimistic, a larger wet share
also gives incentives for the median voter to prefer more law enforcement, since the perceived moral
externality is larger for a given moral view, while the expected cost of increased crime is low.

On the other hand, policing choices would have been much more stable over time because the in-
creased salience of the moral externality reduces the extent to which the police expenditure responds
to changes in the alcohol supply. Nevertheless, as an added equilibrium effect, the distribution of
police enforcement choices across cities spreads out considerably. The apparent reason is a political
economy effect; because a larger wet share shifts the median voter towards “wetness”, there is a force
driving the equilibrium choice of law enforcement downwards. Finally, the model predicts that these
polarized communities would observe significantly higher levels of crime during Prohibition. For
instance, the median city would have on average 2.9 more homicides per hundred thousand on the
average Prohibition year in the two standard deviations higher polarization society, or 1.37 more in
the one standard deviation higher polarization case. Thus, although communities with more extreme
preference distributions do learn much faster about the structural relationship between Prohibition
and crime, they also face a constituency much more willing to endure the increased levels of crime.

6.4.4 Alternative Political Environments

In the setting of this model, it is natural to ask what would the equilibrium effects of changes in
the political environment be. This is important because, as I have shown, the equilibrium collective
law enforcement decisions play a central role in the success or failure of a given legal standard. In
particular, I ask about the effect of interest groups in politics by assuming that some constituencies
have more political power than others, shifting the decisive voter away from the median. To make
the intuitions clear I look at the polar cases in which the decisive voter in the community is either
the median voter among the wets (the decisive voter’s type is \( q_j = -1 \)), or the median voter among
the drys (the decisive voter’s type is \( q_j = 0 \)). Under each counterfactual scenario I compute the
predicted outcome sequences, using the benchmark parameter estimates.

Results are closely related to the ones above. When drys have all the political power, law en-
forcement choices are consistently larger in magnitude relative to the benchmark case. Because
alcohol demand remains unchanged, these enforcement choices increase the informativeness of the
crime signals, making beliefs evolve faster. Belief sequences across the distribution of cities under
this counterfactual scenario are on average two years ahead relative to the benchmark case. Consis-
tently, the predictions of the counterfactual simulation where wets have all the political power
deliver weaker law enforcement relative to the benchmark, which consequently translates into slower
learning. The benchmark estimated average beliefs in 1925 would only be reached in 1929 under
this counterfactual setting.

These results are driven by the increased divergence between the decisive voter over law enforcement
and average voters’ preferences. They make the point that the effects of increased conflict also arise
when the identity of those deciding over law enforcement is further away from overall constituency preferences. Increased conflict, in this setting due to a skewed collective decisionmaking process, is a force driving changes in public opinion. When drys have all the political power at the local level, their choices of law enforcement are too large relative to what the community’s median voter would prefer, and relative to the community’s alcohol market size. As a result, crime outcomes are more informative and communities learn faster. In the polar opposite case, when all political power is allocated to the wets, learning is too slow relative to the benchmark because the very weak law enforcement choices make crime realizations uninformative.

7 Conclusions

Many central political cleavages in contemporary societies revolve around ideological or moral issues, over which people frequently have strong and polarized views. I have highlighted learning about policies, and the endogenous dynamic feedback between enforcement choices and policy support, as a driving force for changes in public opinion over moral issues, and more broadly for social change, by looking at the U.S. Prohibition experience during the early decades of the Twentieth century. The circumstances around Prohibition were very specific to that policy; in particular, the potential effects that closing the alcohol market could have over crime are very specific to prohibitions. Nevertheless, looking at the side-effects (or absence thereof) of policies, and at learning about them, can allow a better understanding of the evolution of policy reform over social cleavages. The extent to which people are informed is important, and of course, the political economy of the extent of such information acquisition becomes key; this should be an area of future research.

I developed a model of endogenous learning and law enforcement in a political economy framework, which has some success in replicating the patterns observed in the data. The paper suggests that an important element to understand the effects and success of policies is the degree of alignment of the legal standard and the law enforcement choices associated with it. This was particularly relevant during Prohibition because most of the law enforcement was decided at the local level, while the prohibitionist legal standard was chosen either at the state or nationwide levels. The estimates suggests that prior beliefs about Prohibition’s effect on crime were very optimistic and highly correlated with moral views, that local policy responded closely to communities’ preferences, and that community preferences also were responsive to changes in beliefs. In the model, the estimated speed of learning is relatively slow. This might be due to the assumption of exclusively localized learning, whereas it is likely that individuals’ opinions were also affected by outcomes across the country. Learning from neighboring communities is likely to be important in societies where the media plays a large role in shaping public opinion. This constitutes an avenue for improvement of the structural model, and for understanding other instances of social change. This paper did not exploit the judicial dimension of law enforcement either, although prohibition enforcement at the local level was also implemented through judicial prosecution. Further research should look at the evolution of judicial decision-making regarding Prohibition as an alternative law enforcement mechanism, which was likely subject to different political economy incentives.
References


League, Anti-Saloon, Anti-Saloon League Yearbook, Anti-Saloon League of America, 1932.


Lewis, Michael, “Accounting for Differences in Local and State Alcohol Laws, North Carolina in 1908,” NA.


# Tables

## Table 1: Prohibition Enforcement during the 1920s

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Midwest</td>
<td>Seized Distilleries, Stills and Still Worms</td>
<td>8,058</td>
<td>9,171</td>
<td>10,207</td>
<td>11,369</td>
<td>5,095</td>
</tr>
<tr>
<td></td>
<td>Seized Fermenters</td>
<td>6,830</td>
<td>7,525</td>
<td>48,748</td>
<td>81,178</td>
<td>8,714</td>
</tr>
<tr>
<td></td>
<td>Seized Spirits*</td>
<td>1,050,017</td>
<td>610,440</td>
<td>472,922</td>
<td>475,840</td>
<td>505,713</td>
</tr>
<tr>
<td></td>
<td>Seized Malt, Wine, Cider, Mash and Pomace*</td>
<td>3,675,499</td>
<td>7,034,847</td>
<td>11,206,587</td>
<td>15,029,002</td>
<td>15,915,534</td>
</tr>
<tr>
<td></td>
<td>Seized Autos and Boats</td>
<td>1,095</td>
<td>1,873</td>
<td>2,949</td>
<td>3,069</td>
<td>4,154</td>
</tr>
<tr>
<td></td>
<td>Killed or Injured Officers</td>
<td>14</td>
<td>15</td>
<td>38</td>
<td>48</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Federal Arrests</td>
<td>24,150</td>
<td>28,185</td>
<td>31,755</td>
<td>25,528</td>
<td></td>
</tr>
<tr>
<td></td>
<td>State Arrests</td>
<td>8,335</td>
<td>7,500</td>
<td>8,714</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>Seized Distilleries, Stills and Still Worms</td>
<td>4,191</td>
<td>3,456</td>
<td>11,136</td>
<td>6,960</td>
<td>4,511</td>
</tr>
<tr>
<td></td>
<td>Seized Fermenters</td>
<td>2,506</td>
<td>3,411</td>
<td>50,079</td>
<td>53,973</td>
<td>9,264</td>
</tr>
<tr>
<td></td>
<td>Seized Spirits*</td>
<td>206,411</td>
<td>767,086</td>
<td>1,142,467</td>
<td>929,877</td>
<td>1,753,629</td>
</tr>
<tr>
<td></td>
<td>Seized Malt, Wine, Cider, Mash and Pomace*</td>
<td>1,143,955</td>
<td>6,334,026</td>
<td>11,811,643</td>
<td>13,930,121</td>
<td>21,233,629</td>
</tr>
<tr>
<td></td>
<td>Seized Autos and Boats</td>
<td>1,082</td>
<td>4,078</td>
<td>4,333</td>
<td>3,334</td>
<td>4,115</td>
</tr>
<tr>
<td></td>
<td>Killed or Injured Officers</td>
<td>4</td>
<td>24</td>
<td>26</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Federal Arrests</td>
<td>35,316</td>
<td>46,396</td>
<td>29,657</td>
<td>47,585</td>
<td></td>
</tr>
<tr>
<td></td>
<td>State Arrests</td>
<td>3,610</td>
<td>3,828</td>
<td>3,708</td>
<td>2,848</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>Seized Distilleries, Stills and Still Worms</td>
<td>34,087</td>
<td>48,038</td>
<td>43,327</td>
<td>43,273</td>
<td>31,671</td>
</tr>
<tr>
<td></td>
<td>Seized Fermenters</td>
<td>179,280</td>
<td>238,528</td>
<td>267,605</td>
<td>301,521</td>
<td>8,513</td>
</tr>
<tr>
<td></td>
<td>Seized Spirits*</td>
<td>284,888</td>
<td>732,713</td>
<td>647,875</td>
<td>799,950</td>
<td>842,375</td>
</tr>
<tr>
<td></td>
<td>Seized Malt, Wine, Cider, Mash and Pomace*</td>
<td>11,933,042</td>
<td>22,026,530</td>
<td>30,428,757</td>
<td>34,753,824</td>
<td>29,953,542</td>
</tr>
<tr>
<td></td>
<td>Seized Autos and Boats</td>
<td>2,703</td>
<td>4,979</td>
<td>5,290</td>
<td>7,058</td>
<td>9,054</td>
</tr>
<tr>
<td></td>
<td>Killed or Injured Officers</td>
<td>18</td>
<td>42</td>
<td>62</td>
<td>122</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Federal Arrests</td>
<td>42,673</td>
<td>49,498</td>
<td>53,300</td>
<td>51,976</td>
<td></td>
</tr>
<tr>
<td></td>
<td>State Arrests</td>
<td>11,778</td>
<td>11,001</td>
<td>12,811</td>
<td>7,737</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>Seized Distilleries, Stills and Still Worms</td>
<td>4,691</td>
<td>7,089</td>
<td>6,334</td>
<td>4,956</td>
<td>2,141</td>
</tr>
<tr>
<td></td>
<td>Seized Fermenters</td>
<td>4,855</td>
<td>14,463</td>
<td>21,291</td>
<td>18,711</td>
<td>1,745</td>
</tr>
<tr>
<td></td>
<td>Seized Spirits*</td>
<td>115,377</td>
<td>230,132</td>
<td>228,029</td>
<td>269,567</td>
<td>283,214</td>
</tr>
<tr>
<td></td>
<td>Seized Malt, Wine, Cider, Mash and Pomace*</td>
<td>1,171,349</td>
<td>3,019,286</td>
<td>3,952,164</td>
<td>7,374,756</td>
<td>5,242,851</td>
</tr>
<tr>
<td></td>
<td>Seized Autos and Boats</td>
<td>1,138</td>
<td>1,421</td>
<td>1,774</td>
<td>2,040</td>
<td>3,009</td>
</tr>
<tr>
<td></td>
<td>Killed or Injured Officers</td>
<td>18</td>
<td>20</td>
<td>31</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Federal Arrests</td>
<td>17,007</td>
<td>14,486</td>
<td>16,845</td>
<td>9,943</td>
<td></td>
</tr>
<tr>
<td></td>
<td>State Arrests</td>
<td>4,718</td>
<td>6,828</td>
<td>6,640</td>
<td>9,262</td>
<td></td>
</tr>
</tbody>
</table>

*Gallons

Northeast includes ME, NH, VT, MA RI, CT, NY, PA, and NJ
Midwest includes ND, SD, NE, KS, MN, IA, MO, WI, IL, IN, MI, OH
South includes DE, MD, DC, VA, WV, KY, NC, TN, SC, GA, AL, MS, FL, AR, LA, OK, and TX
West includes WA, OR, CA, ID, MT, WY, CO, UT, NV, AZ, NM
Table 2: Summary Statistics

<table>
<thead>
<tr>
<th>Summary Statistics</th>
<th>Midwest</th>
<th>Northeast</th>
<th>South</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dry Religions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Baptist (1916)</td>
<td>12.066</td>
<td>8.601</td>
<td>29.660</td>
<td>16.149</td>
</tr>
<tr>
<td>(8.559)</td>
<td>(5.901)</td>
<td>(13.921)</td>
<td>(9.245)</td>
<td></td>
</tr>
<tr>
<td>% Evangelical (1916)</td>
<td>2.623</td>
<td>0.647</td>
<td>0.986</td>
<td>0.852</td>
</tr>
<tr>
<td>(2.083)</td>
<td>(1.393)</td>
<td>(1.496)</td>
<td>(0.510)</td>
<td></td>
</tr>
<tr>
<td>% Mormon (1916)</td>
<td>0.298</td>
<td>0.057</td>
<td>0.160</td>
<td>5.198</td>
</tr>
<tr>
<td>(0.570)</td>
<td>(0.140)</td>
<td>(0.413)</td>
<td>(18.856)</td>
<td></td>
</tr>
<tr>
<td>% Presbyterian (1916)</td>
<td>4.779</td>
<td>4.471</td>
<td>6.060</td>
<td>8.068</td>
</tr>
<tr>
<td>(2.642)</td>
<td>(3.521)</td>
<td>(3.246)</td>
<td>(4.743)</td>
<td></td>
</tr>
<tr>
<td><strong>Wet Religions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Eastern Orthodox (1916)</td>
<td>1.139</td>
<td>1.272</td>
<td>0.437</td>
<td>1.890</td>
</tr>
<tr>
<td>(1.443)</td>
<td>(1.461)</td>
<td>(0.591)</td>
<td>(2.222)</td>
<td></td>
</tr>
<tr>
<td>% Jewish (1916)</td>
<td>1.741</td>
<td>2.771</td>
<td>1.716</td>
<td>1.887</td>
</tr>
<tr>
<td>(1.130)</td>
<td>(2.515)</td>
<td>(1.303)</td>
<td>(1.033)</td>
<td></td>
</tr>
<tr>
<td>% Lutheran (1916)</td>
<td>7.677</td>
<td>3.510</td>
<td>2.521</td>
<td>3.181</td>
</tr>
<tr>
<td>(5.708)</td>
<td>(3.678)</td>
<td>(2.119)</td>
<td>(1.944)</td>
<td></td>
</tr>
<tr>
<td>% Catholic (1916)</td>
<td>55.798</td>
<td>66.631</td>
<td>31.496</td>
<td>47.494</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Black (1910)</td>
<td>3.104</td>
<td>2.325</td>
<td>26.525</td>
<td>1.336</td>
</tr>
<tr>
<td>(2.823)</td>
<td>(2.057)</td>
<td>(12.398)</td>
<td>(0.951)</td>
<td></td>
</tr>
<tr>
<td>% Foreign White (1910)</td>
<td>24.573</td>
<td>32.618</td>
<td>7.928</td>
<td>23.038</td>
</tr>
<tr>
<td>(10.557)</td>
<td>(8.664)</td>
<td>(5.705)</td>
<td>(5.364)</td>
<td></td>
</tr>
<tr>
<td>% Native White (1910)</td>
<td>72.220</td>
<td>64.945</td>
<td>65.412</td>
<td>73.210</td>
</tr>
<tr>
<td>(9.036)</td>
<td>(8.170)</td>
<td>(9.816)</td>
<td>(5.610)</td>
<td></td>
</tr>
<tr>
<td>(0.909)</td>
<td>(1.010)</td>
<td>(1.166)</td>
<td>(0.909)</td>
<td></td>
</tr>
<tr>
<td>% Ages 25-44 (1910)</td>
<td>34.258</td>
<td>33.515</td>
<td>34.159</td>
<td>38.519</td>
</tr>
<tr>
<td>(1.610)</td>
<td>(1.204)</td>
<td>(2.163)</td>
<td>(3.087)</td>
<td></td>
</tr>
<tr>
<td><strong>Legislation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Dry Laws (1919)</td>
<td>6.709</td>
<td>3.376</td>
<td>5.999</td>
<td>9.303</td>
</tr>
<tr>
<td>(3.136)</td>
<td>(2.014)</td>
<td>(1.696)</td>
<td>(3.310)</td>
<td></td>
</tr>
<tr>
<td>Number of Years Under Prohibition**</td>
<td>15.925</td>
<td>15.018</td>
<td>18.035</td>
<td>15.844</td>
</tr>
<tr>
<td>(4.443)</td>
<td>(4.080)</td>
<td>(1.253)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per Capita Police Expenditure (1913 prices)</td>
<td>1.541</td>
<td>1.793</td>
<td>1.977</td>
<td>1.726</td>
</tr>
<tr>
<td>(0.726)</td>
<td>(0.759)</td>
<td>(0.798)</td>
<td>(0.800)</td>
<td>(0.687)</td>
</tr>
<tr>
<td>Police Expenditure Share</td>
<td>0.108</td>
<td>0.092</td>
<td>0.112</td>
<td>0.112</td>
</tr>
<tr>
<td>(0.032)</td>
<td>(0.026)</td>
<td>(0.021)</td>
<td>(0.024)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Drunkenness Arrest Rate (per 1,000)</td>
<td>16.000</td>
<td>14.459</td>
<td>16.653</td>
<td>12.132</td>
</tr>
<tr>
<td>Homicide Rate (per 100,000)</td>
<td>10.807</td>
<td>18.124</td>
<td>5.368</td>
<td>10.076</td>
</tr>
<tr>
<td>% Anti-Prohibition vote share***</td>
<td>0.518</td>
<td>0.723</td>
<td>0.523</td>
<td>0.826</td>
</tr>
<tr>
<td>(0.169)</td>
<td>(0.168)</td>
<td>(0.132)</td>
<td>(0.129)</td>
<td>(0.0220)</td>
</tr>
</tbody>
</table>

*Regions as as classified by the Bureau of the Census: North East includes ME, NH, VT, MA, RI, CT, NY, PA, and NJ
Midwest includes ND, SD, NE, KS, MN, IA, MO, WI, IL, IN, MI, OH
South includes DE, MD, DC, VA, WV, KY, NC, TN, SC, GA, AL, MS, FL, AR, LA, OK, and TX
West includes WA, OR, CA, ID, MT, WY, CO, UT, NV, AZ, and NM
** During the 1910-1933 period
*** From state level referenda
Standard Deviations in parenthesis
All summary statistics are weighted by city population
### Table 3: Public Opinion Regressions

#### Panel A: Electoral Support for Prohibition

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Sample</th>
<th>All Counties</th>
<th>County Sample</th>
<th>Counties with Pop&gt;30,000</th>
<th>City Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-Prohibition Indicator</td>
<td></td>
<td>0.1373</td>
<td>0.1099</td>
<td>0.0806</td>
<td>0.1714</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.018)</td>
<td>(0.022)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>&quot;Wetness&quot;</td>
<td></td>
<td>0.1201</td>
<td>0.1422</td>
<td>0.1211</td>
<td>0.1073</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.090)</td>
<td>(0.095)</td>
<td>(0.078)</td>
<td>(0.096)</td>
</tr>
<tr>
<td>Baseline &quot;Wetness&quot; x Post-Prohibition Indicator</td>
<td></td>
<td>0.6969</td>
<td>0.5985</td>
<td>0.4808</td>
<td>0.7349</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.045)</td>
<td>(0.042)</td>
<td>(0.044)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>log of Population</td>
<td></td>
<td>0.0881</td>
<td>0.0671</td>
<td>0.0806</td>
<td>0.0821</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Urban share of county</td>
<td></td>
<td>0.0213</td>
<td>0.0517</td>
<td>0.0572</td>
<td>0.0127</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.039)</td>
<td>(0.028)</td>
<td>(0.041)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Number of Dry Laws</td>
<td></td>
<td>-0.0109</td>
<td>-0.0102</td>
<td>-0.0546</td>
<td>-0.0342</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.031)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Referendum Year</td>
<td></td>
<td>-0.0048</td>
<td>0.0072</td>
<td>0.0024</td>
<td>-0.0019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.014)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Referendum Type:</td>
<td></td>
<td>-0.0048</td>
<td>0.0072</td>
<td>0.0024</td>
<td>-0.0019</td>
</tr>
<tr>
<td>Prohibition Law</td>
<td></td>
<td>0.0100</td>
<td>0.0386</td>
<td>0.1141</td>
<td>-0.0022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.016)</td>
<td>(0.045)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Inverse Mills Ratio x Pre-prohibition Period</td>
<td></td>
<td>0.1577</td>
<td>0.1896</td>
<td>-0.0546</td>
<td>-0.0546</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.031)</td>
<td>(0.053)</td>
<td>(0.073)</td>
<td>(0.073)</td>
</tr>
</tbody>
</table>

#### Panel B: Probit Selection Equation

<table>
<thead>
<tr>
<th>State cross Post-Prohibition Effects</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>R squared</td>
<td>0.584</td>
<td>0.488</td>
<td>0.555</td>
<td>0.766</td>
<td>0.525</td>
<td>0.498</td>
<td>0.644</td>
<td>0.688</td>
<td>0.876</td>
<td>0.665</td>
<td>0.741</td>
<td>0.774</td>
<td>0.613</td>
<td>0.943</td>
<td>0.792</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of Cross Sections</td>
<td>1983</td>
<td>1693</td>
<td>1693</td>
<td>1693</td>
<td>1693</td>
<td>1693</td>
<td>672</td>
<td>672</td>
<td>672</td>
<td>672</td>
<td>258</td>
<td>258</td>
<td>258</td>
<td>258</td>
<td>258</td>
<td>258</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of Observations</td>
<td>2435</td>
<td>2386</td>
<td>2386</td>
<td>2386</td>
<td>2386</td>
<td>2386</td>
<td>672</td>
<td>672</td>
<td>672</td>
<td>672</td>
<td>258</td>
<td>258</td>
<td>258</td>
<td>258</td>
<td>258</td>
<td>258</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Note:
- Constant for the Vote share equations not reported. Standard Errors are robust and clustered at the county or city level.
- The Selection equation is a probit at the state level, of an indicator for having had an Alcohol referendum in the Pre-Prohibition period, on the state-level share of adherents to any "wet" religion (Orthodox, Jewish, Luther, Catholic, Other) from the 1916 Census of Religions, and the share of non-native white individuals in the population, from the 1910 Census of Population.
Table 4: Conditional Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficients</th>
<th>Conditional Maximum Likelihood Estimates</th>
<th>Covariate</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>a % ages 15-44*</td>
<td>0.606 (0.273)</td>
<td>a Constant 2.758 (0.762)</td>
<td>B % Baptist in 1911</td>
<td>-0.237 (0.071)</td>
</tr>
<tr>
<td>% Foreign White*</td>
<td>0.519 (0.198)</td>
<td>Number of Alcohol-Related Laws 0.117 (0.248)</td>
<td>% Orthodox in 1911</td>
<td>0.022 (0.020)</td>
</tr>
<tr>
<td>% Black*</td>
<td>0.199 (0.058)</td>
<td>Enforcement Law -0.895 (0.206)</td>
<td>% Evangelical in 1911</td>
<td>-0.034 (0.008)</td>
</tr>
<tr>
<td>% Baptist*</td>
<td>0.382 (0.281)</td>
<td>Prohibition Unit Seat -0.227 (0.153)</td>
<td>% Jewish in 1911</td>
<td>-0.021 (0.031)</td>
</tr>
<tr>
<td>% Orthodox*</td>
<td>0.434 (0.223)</td>
<td>Prohibition Unit: Providence -0.281 (0.201)</td>
<td>% Mormon in 1911</td>
<td>-0.019 (0.677)</td>
</tr>
<tr>
<td>% Evangelical*</td>
<td>0.589 (0.498)</td>
<td>Prohibition Unit: Washington -0.909 (0.496)</td>
<td>% Lutheran in 1911</td>
<td>-0.116 (0.078)</td>
</tr>
<tr>
<td>% Jewish*</td>
<td>-0.447 (0.462)</td>
<td>Prohibition Unit: Jacksonville -1.090 (1.432)</td>
<td>% Methodist in 1911</td>
<td>-0.289 (0.161)</td>
</tr>
<tr>
<td>% Mormon*</td>
<td>0.257 (0.259)</td>
<td>Prohibition Unit: Detroit -0.851 (0.326)</td>
<td>% Catholic in 1911</td>
<td>-0.738 (0.019)</td>
</tr>
<tr>
<td>% Lutheran*</td>
<td>0.767 (0.295)</td>
<td>Prohibition Unit: Chicago -2.065 (1.392)</td>
<td>% Presbyterian in 1911</td>
<td>-0.034 (0.011)</td>
</tr>
<tr>
<td>% Methodist*</td>
<td>0.746 (0.555)</td>
<td>Prohibition Unit: Kansas City -1.166 (0.725)</td>
<td>Constant</td>
<td>-2.231 (0.612)</td>
</tr>
<tr>
<td>% Catholic*</td>
<td>0.174 (0.041)</td>
<td>Prohibition Unit: San Francisco 2.125 (1.217)</td>
<td>c</td>
<td>0.277 (0.084)</td>
</tr>
<tr>
<td>% Presbyterian*</td>
<td>-0.255 (0.016)</td>
<td>Prohibition Unit: Los Angeles 0.004 (0.603)</td>
<td>c</td>
<td>0.061 (0.028)</td>
</tr>
<tr>
<td>Log of Population*</td>
<td>0.001 (0.034)</td>
<td>Prohibition Unit: Seattle 1.669 (0.659)</td>
<td>c</td>
<td>0.348 (0.135)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.477 (0.088)</td>
<td>z % Baptist 0.890 (0.395)</td>
<td>c</td>
<td>1.000 (0.000)</td>
</tr>
<tr>
<td>b</td>
<td>8.666 (2.773)</td>
<td>% Orthodox -0.223 (0.076)</td>
<td>p</td>
<td>-0.495 (0.089)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Evangelical 0.187 (0.082)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Jewish 0.728 (0.650)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Mormon 0.172 (0.212)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Lutheran 0.071 (0.296)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Methodist 1.186 (0.438)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Catholic 0.819 (0.669)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Presbyterian -0.772 (1.370)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constant 0.024 (0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Presbyterian 0.073 (0.059)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Baptist 0.231 (0.088)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constant 1.504 (0.713)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Log-likelihood: 6650.774
Observations: 1254

*1911-1929 averages
Note: Standard Errors are computed through a bootstrap of size 100.
Estimates of the State Effects are omitted from the table.
<table>
<thead>
<tr>
<th>Model Parameters</th>
<th>Mean Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>3.167</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
</tr>
<tr>
<td>b</td>
<td>8.666</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>( \chi )</td>
<td>8.948</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.266</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.259</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>3.437</td>
</tr>
<tr>
<td></td>
<td>(1.010)</td>
</tr>
<tr>
<td>( z )</td>
<td>0.687</td>
</tr>
<tr>
<td></td>
<td>(0.108)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>1.377</td>
</tr>
<tr>
<td></td>
<td>(0.258)</td>
</tr>
<tr>
<td>( B )</td>
<td>-1.311</td>
</tr>
<tr>
<td></td>
<td>(0.265)</td>
</tr>
</tbody>
</table>

Note: Standard Deviations in parenthesis
Figures

Figure 1: The Homicide Rate in U.S. Cities, 1911-1936 (per 100,000)\textsuperscript{42}

Figure 2: The Drunkenness Arrest Rate in U.S. Cities, 1911-1929 (per capita)

\textsuperscript{42}In all figures, red lines are 95\% confidence intervals.
Figure 3: Timing of State Adoption of Prohibition
Figure 4: $\delta_r$'s from equation (1):

Panel A: Homicide Rate (per 100,000) (left) and Drunkenness Arrest Rate (per 1,000) (right)

Panel B: Police Expenditure Share (left) and Per Capita Police Expenditure (right)
Figure 5: $\phi_s$'s from Equation (3):

Panel A: Homicide Rate (per 100,000) (left) and Drunkenness Arrest Rate (per 1,000) (right)

Panel B: Police Expenditure Share (left) and Per Capita Police Expenditure (right)
Figure 6: Alcohol Referenda and Public Opinion shift

Figure 7: Moral Views and Changes in Public Opinion (U.S. counties)
Figure 8: Fit of the Model

Panel A: Cross-Sectional Fit

Panel B: Time-Series Fit
Figure 9: Distribution of the Estimated $\theta_e$'s

Figure 10: Estimated Prior Beliefs $B_e$ vs. Estimated Average Moral views $\bar{z}_e$. 
Figure 11: Estimated Functional Forms

Alcohol Supply Catch-up

Drunk Arrest Probability

Prohibition-related % Increase in Homicide Rate

\[ \kappa = 0.26, \lambda = 0.25 \]

\[ \chi = 0.94, \alpha = 3.43 \]

\[ \theta = 1.37, \alpha = 3.43, E[u] = 0.267, K(t) = 0.8 \]

Figure 12: Estimated Densities of the Median Voters’ Unobserved \( q^{med} \)

\[ a = 3.5, b = 8.6, \alpha_1 = 0.34, \alpha_2 = 1, \rho = 0.49, \Omega = 0.34 \]

(No Prohibition: Purple, Prohibition: Pink)
Figure 13: Estimated Belief Sequences: Empirical Distribution
Figure 14: "Out of Sample" predictions for the years 1930-1936

Panel A: Cross-Sectional Fit

Panel B: Time-Series Fit
Appendix 1: Derivation of Ideal Law Enforcement Choice

Indirect utility under no Prohibition is given by

\[ E_t U_t(p_t|P_t = 0) = 1_{\{i \in W_t\}} \exp(-\alpha_t p_t) - \frac{z^i}{a + b} \exp(-\alpha_t p_t) + \exp(1 - p_t) - \frac{a}{a + b} \exp(-\alpha_t p_t) \]  
(26)

The first order condition with respect to \( p_{ct} \) from equation (26) is

\[ -1_{\{i \in W_t\}} \alpha_c(\frac{a}{a + b} - \frac{z^i}{k(T_t)} - \frac{\alpha_t}{a + b} \exp(-\alpha_t p_{ct})) \leq 0 \]

Solving for \( p_{ct} \), equation (16) directly follows. The second order condition for this problem is given by

\[ 1_{\{i \in W_t\}} \frac{2}{\alpha c} \exp(-\alpha_t p_{ct}) - \frac{2}{\alpha c} \frac{a}{a + b} \exp(-\alpha_t p_{ct}) + \exp(1 - p_{ct}) - \frac{2}{\alpha c} \frac{a}{a + b} \exp(-\alpha_t p_{ct}) < 0 \]

\[ \Rightarrow 2ln\alpha_c + ln \left[ \frac{a_c}{a_c + b} (z_c^i + 1) - 1_{\{i \in W_t\}} \right] - 1 > (\alpha_c - 1)p_{ct} \]  
(27)

I verify this condition is satisfied for the parameter estimates.

Indirect utility under Prohibition is given by

\[ E_t U_t(p_t|P_t = 1) = 1_{\{i \in W_t\}} k(T_t) \exp(-\alpha_t p_t) - \frac{z^i}{a + b} k(T_t) \exp(-\alpha_t p_t) + \exp(1 - p_t) - \frac{a}{a + b} k(T_t) \exp(-\alpha_t p_t) \]

\[ + \exp(1 - p_t) - \frac{a}{a + b} k(T_t) \exp(-\alpha_t p_t) - \frac{\alpha_t}{a + b} [1 - \exp(-\alpha_t p_t)] \]  
(28)

The first order condition with respect to \( p_{ct} \) from equation (28) is

\[ -1_{\{i \in W_t\}} k(T_{ct}) \alpha_c(\frac{a}{a + b} - \frac{z^i}{k(T_{ct})} - \frac{\alpha_t}{a + b} \exp(-\alpha_t p_{ct})) \leq 0 \]

Solving for \( p_{ct} \), equation (17) directly follows. The second-order condition for the solution in equation (17) to be a maximum is

\[ \Rightarrow 2ln\alpha_c + ln \left[ \frac{a_c}{a_c + b} (z_c^i - \frac{\alpha_t}{a + b} + 1) - 1_{\{i \in W_t\}} \right] - 1 > (\alpha_c - 1)p_{ct} \]  
(29)
Appendix 2: Proof of Proposition 1

Proof. In this community there are three sources of heterogeneity in preferences over law enforcement: the distribution of moral views, the distribution of belief biases, and the distribution of types (wet and dry). First, observe that conditional on \((\xi, \xi')\), the preferred level of law enforcement of a wet voter is shifted down by a constant factor relative to the optimal choice of a dry individual. Thus, for periods under Prohibition define \(\hat{\theta}_\text{DP}^i = \frac{a}{a+b}(\xi - \Omega t \frac{1}{\sigma^2_\xi}) (DP \text{ for Dry under Prohibition})\), and \(\hat{\theta}_\text{WP}^i = \frac{a}{a+b}(\xi - \Omega t \frac{1}{\sigma^2_\xi}) - 1 (WP \text{ for Wet under Prohibition})\). These are normal random variables distributed according to \(\hat{\theta}_\text{DP}^i \sim N(0, \sigma^2_{\theta_{\text{WP}^i}})\) and \(\hat{\theta}_\text{WP}^i \sim N(-1, \sigma^2_{\theta_{\text{WP}^i}})\) respectively, where \(\sigma^2_{\theta_{\text{WP}^i}} \equiv \left(\frac{a}{a+b}\right)^2 \left(\sigma^2_\xi + \Omega^2 t \frac{1}{\sigma^2_\xi} - 2\Omega t \frac{\sigma_\xi}{\sigma^2_\xi}\right)^2\). Now define \(\hat{\theta}_P^i = 1_{\{i \in D_t\}} \hat{\theta}_\text{DP}^i + 1_{\{i \in W_t\}} \hat{\theta}_\text{WP}^i\). The conditional density of \(\hat{\theta}_P^i\) is given by

\[
 f_\theta(\hat{\theta}_P^i | \mu_t) = (1 - \mu_t)N(0, \sigma^2_{\theta_{\text{WP}^i}}) + \mu_t N(-1, \sigma^2_{\theta_{\text{WP}^i}})
\]

since with probability \(\mu_t\) a wet individual is sampled, and with probability \(1 - \mu_t\) a dry individual is sampled. Thus, the distribution of \(\hat{\theta}_P^i\) in the population is a mixture of two normal random variables with a common variance, one of which is shifted to the left by 1 relative to the other. Given the normality of \(\hat{\theta}_\text{WP}^i\) and \(\hat{\theta}_\text{DP}^i\), as \(\mu_t \to 0\), the median voter’s type \(\theta_{\text{med}}^i \to 0\), and as \(\mu_t \to 1\), \(\theta_{\text{med}}^i \to -1\), so that \(\theta_{\text{med}}^i \in (-1, 0)\). For periods under no Prohibition, analogously define \(\hat{\theta}_\text{DN}^i \equiv \frac{a}{a+b}\xi (DN \text{ for dry under no Prohibition})\) and \(\hat{\theta}_\text{WN}^i \equiv \frac{a}{a+b}\xi - 1 (WN \text{ for wet under no Prohibition})\), which are distributed according to \(\hat{\theta}_\text{DN}^i \sim N(0, \sigma^2_{\theta_{\text{WN}^i}})\) and \(\hat{\theta}_\text{WN}^i \sim N(-1, \sigma^2_{\theta_{\text{WN}^i}})\) respectively, with \(\sigma^2_{\theta_{\text{WN}^i}} \equiv \left(\frac{a}{a+b}\right)^2 \sigma^2_\xi\). Now define \(\hat{\theta}_N^i = 1_{\{i \in D_t\}} \hat{\theta}_\text{DN}^i + 1_{\{i \in W_t\}} \hat{\theta}_\text{WN}^i\), which is a random variable whose conditional density is given by

\[
 f_\theta(\hat{\theta}_N^i | \mu_t) = (1 - \mu_t)N(0, \sigma^2_{\theta_{\text{WN}^i}}) + \mu_t N(-1, \sigma^2_{\theta_{\text{WN}^i}})
\]

Indirect preferences over law enforcement in (16) and (17) can be expressed in terms of \(\theta_N^i\) and \(\theta_P^i\). It follows that this is a purely private-values election because individuals realize that differences in beliefs are due to individual-specific biases. For a given individual, the voting decisions of the members of his community do not convey any additional information. Moreover, indirect preferences over law enforcement are single-peaked in \(\theta_P^i\), so the Median Voter Theorem holds, and the unique political equilibrium value of \(\mu_t\) is given by the preferred choice of law enforcement of the median over the distribution of \(\theta_P^i\), conditional on \(\mu_t\).

The (conditional) median for Prohibition years will be given by the value of \(\theta_{\text{med}}^i\) which solves the following equation

\[
 (1 - \mu_t) \int_{-\infty}^{\theta_{\text{med}}^i} \frac{1}{\sqrt{2\pi\sigma_{\theta_P}}} \exp \left( -\frac{1}{2\sigma^2_{\theta_P}} \theta^2 \right) d\theta + \mu_t \int_{-\infty}^{\theta_{\text{med}}^i} \frac{1}{\sqrt{2\pi\sigma_{\theta_P}}} \exp \left( -\frac{1}{2\sigma^2_{\theta_P}} (\theta + 1)^2 \right) d\theta = \frac{1}{2}
\]

where I have made explicit the dependence of \(\theta_{\text{med}}^i\) on the wet share in the community. Because the realization of \(\mu_t\) is unobserved, the median \(\theta_{\text{med}}^i\) in the population as defined in (30) is a random

\[43\] This variance is time-varying. As learning takes place and \(\Omega_t \to 0\), \(\sigma^2_{\theta_P} \to \sigma^2_\xi\).
variable whose density is derived below. The equation analogous to (30) implicitly defining $g_{med}^N$ (the conditional median of the distribution of $g_N^t$) and its corresponding density are found analogously.\footnote{Notice that $\sigma^2_{\text{ct}} \to \sigma^2_{\text{vN}}$ as $\Omega_t \to 0$, which implies that $g_{med}^N \to g_{med}^\text{ct}$.}

Derivation of the density of $g_{med}^p$:

First, recall that $f_{\mu}(\mu; a, b)$, the density of $\mu_{ct}$, is beta with parameters $(a_c, b)$. From (30), $\mu_{ct}$ can be directly expressed as a function of $g_{med}^p$:

$$
\mu_{ct} \equiv h_{\mu}(g_{med}^p) = \frac{\frac{1}{2} - \Phi \left( g_{med}^p \right)}{\Phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right) - \Phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right)}
$$

If this is a one-to-one mapping, the density of $g_{med}^p$ will be given by

$$
f_{g_{med}^p}(g_{med}^p; a_c, b_c, \sigma_{\text{gpt}}) = f_{\mu}(h_{\mu}(g_{med}^p); a_c, b_c) \left| \frac{\partial h_{\mu}(g_{med}^p)}{\partial g_{med}^p} \right|
$$

The derivative of $h_{\mu}$ is given by

$$
\frac{\partial h_{\mu}(g_{med}^p)}{\partial g_{med}^p} = \frac{\frac{1}{\sigma_{\text{gpt}}} \phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \left[ \frac{1}{2} - \Phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right) \right] - \frac{1}{\sigma_{\text{gpt}}} \phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \left[ \frac{1}{2} - \Phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \right]}{\left[ \Phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right) - \Phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \right]^2} < 0
$$

To see that $\frac{\partial h_{\mu}(g_{med}^p)}{\partial g_{med}^p} < 0$ notice that the first term in square brackets is always smaller than the second term in square brackets. For $g_{med}^p \geq 0$, $\phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \geq \phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right)$, and the first term in brackets is more negative than the second term in brackets, so the numerator is negative. For $g_{med}^p < -\frac{1}{2}$, $\phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) < \phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right)$, the second term in brackets is strictly positive, and the first term in brackets is also positive (but smaller than the second term in brackets), so the numerator is negative. For $g_{med}^p \in (-\frac{1}{2}, 0)$, $\phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \geq \phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right)$, the first term in brackets is negative, and the second term in brackets is positive, so the numerator is negative.

Thus, $h_{\mu}$ is a one-to-one mapping, and the likelihood for $g_{med}^p$ is

$$
f_{g_{med}^p}(g_{med}^p; a_c, b, \sigma_{\text{gpt}}) = \frac{1}{\sigma_{\text{gpt}}} \int_0^1 v^{a_c-1}(1-v)^{b-1}dv \times
$$

$$
\phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right) \left[ \frac{1}{2} - \Phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \right]^{a_c} \left[ \frac{1}{2} - \Phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right) \right] \left[ \frac{1}{2} - \Phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \right]^{a_c-1} \left[ \frac{1}{2} - \Phi \left( \frac{g_{med}^p + 1}{\sigma_{\text{gpt}}} \right) \right] \left[ \frac{1}{2} - \Phi \left( \frac{g_{med}^p}{\sigma_{\text{gpt}}} \right) \right]^{a_c+b}
$$

for $g_{med} \in (-1, 0)$, and where $\sigma_{\text{gpt}} = \frac{\sigma_c}{a_t+b} \sqrt{\frac{2}{\sigma_{\text{ct}}^2 + \Omega_c^2 L_{\text{ct}}^2 + 2\Omega_c^2 \Omega_{\text{ct}} \sigma_{\text{ct}}^2}}$.\footnote{Notice that $\sigma^2_{\text{ct}} \to \sigma^2_{\text{vN}}$ as $\Omega_t \to 0$, which implies that $g_{med}^N \to g_{med}^\text{ct}$.}
Replacing $\sigma_{\psi N} = \frac{a_{c} + b_{c}}{a_{c} + b_{c}} \sigma_{\psi}$ for $\sigma_{\psi N}$ everywhere in (32), the density of unobserved heterogeneity in preferred law enforcement during periods under no Prohibition is obtained: $f_{Q^{med}}(Q^{med}; a_{c}, b_{c}, \sigma_{\psi N})$.

Appendix 3: Derivation of the Conditional Likelihood

The joint density function of $(z_{ct}, \mu_{ct}, \epsilon_{ct})$ is given by

$$f_{z_{ct}}(z_{ct}, \mu_{ct}, \epsilon_{ct}; a_{c}, b_{c}, \epsilon_{ct}, k, \lambda, \sigma_{q}^2, \sigma_{z}^2) = \frac{1}{\sqrt{2\pi}\sigma_{z}} \exp \left( -\frac{1}{2\sigma_{z}^2} (z_{ct} - \bar{z}_{ct})^2 \right) \frac{\mu_{ct}^{a_{c}-1}(1 - \mu_{ct})^{b_{c}-1}}{\int \mu_{ct}^{a_{c}-1}(1 - x)^{b_{c}-1}dx \sqrt{2\pi}\sigma_{q}} \exp \left( -\frac{\epsilon_{ct}^2}{2\sigma_{q}^2} \right)$$

From (19), (20), and (21), $(z_{ct}, \mu_{ct}, \epsilon_{ct})$ can be expressed as a function of the observables $(p_{ct}, d_{ct}, q_{ct})$:

From (21),

$$z_{ct} \equiv g_{z}(p_{ct}, d_{ct}, q_{ct}; \epsilon_{ct}, \mu_{ct}, \delta_{ct}) = \frac{a_{c} + b_{c}}{a_{c}} \frac{1}{\alpha_{c}k(\tau_{ct})} \exp(\alpha_{ct}(1)p_{ct} + 1) - \frac{a_{c} + b_{c}}{a_{c}} \left[ P_{ct}\mu_{ct}^{med} + (1 - P_{ct})\gamma_{ct}^{med}\right] + P_{ct}\Omega_{ct}\delta_{ct}^{C} - 1$$

From (20),

$$\mu_{ct} \equiv g_{\mu}(p_{ct}, d_{ct}, q_{ct}) = \frac{d_{ct}}{k(\tau_{ct})} (\chi + \exp(\alpha_{ct}p_{ct}))$$

Finally from (19), and replacing for $\mu_{ct}$ from above,

$$\epsilon_{ct} \equiv g_{\epsilon}(p_{ct}, d_{ct}, q_{ct}) = q_{ct} - \Theta_{S} - d_{ct}(\chi + \exp(\alpha_{ct}p_{ct})) \{ \exp(-\alpha_{ct}p_{ct}) + P_{ct}\theta_{ct}\} [1 - \exp(-\alpha_{ct}p_{ct})] \}$$

Now, if $g(p_{ct}, d_{ct}, q_{ct}) = (g_{z}, g_{\mu}, g_{\epsilon})$ is a one-to-one mapping from $(p_{ct}, d_{ct}, q_{ct})$ to $(z_{ct}, \mu_{ct}, \epsilon_{ct})$, the density function for $(p_{ct}, d_{ct}, q_{ct})$ will be given by

$$f_{pdq}(p_{ct}, d_{ct}, q_{ct}) = f_{z_{ct}}(g_{z}(p_{ct}, d_{ct}, q_{ct}; \epsilon_{ct}, \mu_{ct}, \delta_{ct}), g_{\mu}(p_{ct}, d_{ct}, q_{ct}), g_{\epsilon}(p_{ct}, d_{ct}, q_{ct}); a_{c}, b_{c}, \epsilon_{ct}, k, \lambda, \sigma_{q}^2, \sigma_{z}^2) \left| J_{ct} \right|$$

where $\left| J_{ct} \right|$ is the absolute value of the determinant of the jacobian of $g$:

$$\left| J_{ct} \right| = \begin{vmatrix} \frac{\partial g_{z}}{\partial d} & \frac{\partial g_{z}}{\partial q} & \frac{\partial g_{z}}{\partial \epsilon} \\ \frac{\partial g_{\mu}}{\partial d} & \frac{\partial g_{\mu}}{\partial q} & \frac{\partial g_{\mu}}{\partial \epsilon} \\ \frac{\partial g_{\epsilon}}{\partial d} & \frac{\partial g_{\epsilon}}{\partial q} & \frac{\partial g_{\epsilon}}{\partial \epsilon} \end{vmatrix}$$

Given the structure of the model, conveniently $\frac{\partial g_{z}}{\partial d} = \frac{\partial g_{\mu}}{\partial q} = \frac{\partial g_{\epsilon}}{\partial q} = 0$, and $\frac{\partial g_{\epsilon}}{\partial p} = 1$. To show that $g(p_{ct}, d_{ct}, q_{ct})$ is a one-to-one mapping, it is sufficient that $\frac{\partial g_{z}}{\partial p}, \frac{\partial g_{\mu}}{\partial d}, \frac{\partial g_{\epsilon}}{\partial d}, \text{ and } \frac{\partial g_{\epsilon}}{\partial p}$ do not change sign. Solving for these derivatives.
\[
\frac{\partial g_z}{\partial p} = \frac{a_c + b}{a_c} \alpha_{ct} \frac{\alpha_{ct} - 1}{k(\tau_{ct})} \exp((\alpha_{ct} - 1)p_{ct} + 1)
\]

which is always positive.

\[
\frac{\partial g_{\mu}}{\partial d} = \frac{\chi + \exp(\alpha_{ct}p_{ct})}{k(\tau_{ct})}
\]

which is always positive.

\[
\frac{\partial g_{\mu}}{\partial p} = \frac{d_{ct}}{k(\tau_{ct})} \alpha_{ct} \exp(\alpha_{ct}p_{ct})
\]

which is always positive.

\[
\frac{\partial g_e}{\partial d} = -\left(\chi + \exp(\alpha_{ct}p_{ct})\right) \{\exp(-\alpha_{ct}p_{ct}) + P_{ct} \theta_{ct} [1 - \exp(-\alpha_{ct}p_{ct})]\}
\]

which is always negative. Finally,

\[
\frac{\partial g_e}{\partial p} = -d_{ct} \alpha_{ct} \left[ P_{ct} \theta_{ct} [\exp(\alpha_{ct}p_{ct}) + \chi \exp(-\alpha_{ct}p_{ct})] - \chi \exp(-\alpha_{ct}p_{ct}) \right]
\]

Notice that under no Prohibition, \( \frac{\partial g_e}{\partial p} > 0 \) for any value of \( p_{ct} \). Under Prohibition, a sufficient condition for \( \frac{\partial g_e}{\partial p} < 0 \) (so that total crime is increasing in law enforcement) is that \( \theta_{ct} > \frac{\chi}{\chi + \frac{d_{ct}}{\alpha_{ct}p_{ct}}} \). In this case, \( g(p_{ct}, d_{ct}, q_{ct}) \) is one-to-one, and \(|J| \) reduces to \(|J| = \frac{\partial g_z}{\partial p} \frac{\partial g_{\mu}}{\partial d} \frac{\partial g_e}{\partial q} \).

\[
|J| = \frac{\partial g_z}{\partial p} \frac{\partial g_{\mu}}{\partial d}
\]

**Appendix 4: Additional Reduced Form Results**

In this Appendix I discuss some additional reduced form results. First I present the results of the models whose coefficients are depicted in figures 4 and 5. Then I document the evolution of crime after Prohibition repeal, and look at neighboring alcohol markets and at the effect of pre-Prohibition Dry legislation and Women’s suffrage, as alternative explanations for the main patterns described in section 4. Finally discuss the selection problem in the public opinion models.

**Time-Varying Effects of Prohibition**

Table A4-1 presents the main regression results from estimating equation (1). The tables present estimates of the specification including year effects for each outcome variable. As a benchmark for comparison, the first two columns for each outcome only include an indicator variable for years under Constitutional Prohibition (1920-1933). The next two columns then include the \( D_t \)’s instead
of the Constitutional Prohibition indicator as a way of disaggregating the time-varying effects of Prohibition and allowing for pre-nationwide Prohibition effects.\textsuperscript{46}

Table A4-1: Long and Short-Run effects of Prohibition

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Homicide Rate per 10,000</th>
<th>Drunkenness Arrest Rate per 1,000</th>
<th>Police Expenditure Share</th>
<th>Per Capita Police Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B (3)</td>
<td>C (4)</td>
<td>B (7)</td>
<td>C (8)</td>
</tr>
<tr>
<td>1st Year under Prohibition</td>
<td>-2.290</td>
<td>-2.104</td>
<td>-7.265</td>
<td>-5.593</td>
</tr>
<tr>
<td></td>
<td>(1.417)</td>
<td>(1.389)</td>
<td>(2.146)</td>
<td>(1.287)</td>
</tr>
<tr>
<td>2nd Year under Prohibition</td>
<td>-1.649</td>
<td>-1.252</td>
<td>-10.838</td>
<td>-9.737</td>
</tr>
<tr>
<td></td>
<td>(1.840)</td>
<td>(1.848)</td>
<td>(2.538)</td>
<td>(1.432)</td>
</tr>
<tr>
<td>3rd Year under Prohibition</td>
<td>-1.369</td>
<td>-1.482</td>
<td>-10.947</td>
<td>-9.467</td>
</tr>
<tr>
<td></td>
<td>(1.958)</td>
<td>(1.876)</td>
<td>(2.995)</td>
<td>(1.716)</td>
</tr>
<tr>
<td></td>
<td>(3.071)</td>
<td>(1.978)</td>
<td>(3.600)</td>
<td>(2.008)</td>
</tr>
<tr>
<td>5th Year under Prohibition</td>
<td>-0.072</td>
<td>0.281</td>
<td>-9.095</td>
<td>-7.859</td>
</tr>
<tr>
<td></td>
<td>(2.439)</td>
<td>(2.459)</td>
<td>(3.844)</td>
<td>(2.310)</td>
</tr>
<tr>
<td>6th Year under Prohibition</td>
<td>0.330</td>
<td>0.738</td>
<td>-8.527</td>
<td>-7.422</td>
</tr>
<tr>
<td></td>
<td>(2.818)</td>
<td>(2.785)</td>
<td>(4.216)</td>
<td>(2.623)</td>
</tr>
<tr>
<td>7th Year under Prohibition</td>
<td>0.117</td>
<td>0.731</td>
<td>-9.241</td>
<td>-7.777</td>
</tr>
<tr>
<td></td>
<td>(2.820)</td>
<td>(2.746)</td>
<td>(4.676)</td>
<td>(2.601)</td>
</tr>
<tr>
<td>8th Year under Prohibition</td>
<td>2.381</td>
<td>2.937</td>
<td>-9.238</td>
<td>-8.326</td>
</tr>
<tr>
<td></td>
<td>(2.331)</td>
<td>(2.395)</td>
<td>(4.676)</td>
<td>(2.720)</td>
</tr>
<tr>
<td>9th Year under Prohibition</td>
<td>2.320</td>
<td>3.163</td>
<td>-9.432</td>
<td>-8.083</td>
</tr>
<tr>
<td></td>
<td>(2.628)</td>
<td>(2.709)</td>
<td>(4.996)</td>
<td>(2.801)</td>
</tr>
<tr>
<td>10th Year under Prohibition</td>
<td>1.528</td>
<td>2.397</td>
<td>-9.478</td>
<td>-7.908</td>
</tr>
<tr>
<td></td>
<td>(2.417)</td>
<td>(2.553)</td>
<td>(5.574)</td>
<td>(2.963)</td>
</tr>
<tr>
<td>11th Year under Prohibition</td>
<td>0.841</td>
<td>1.853</td>
<td>-8.045</td>
<td>-7.698</td>
</tr>
<tr>
<td></td>
<td>(2.323)</td>
<td>(2.505)</td>
<td>(5.644)</td>
<td>(2.995)</td>
</tr>
<tr>
<td>12th Year under Prohibition</td>
<td>-0.488</td>
<td>0.904</td>
<td>-5.891</td>
<td>-7.889</td>
</tr>
<tr>
<td></td>
<td>(1.800)</td>
<td>(2.005)</td>
<td>(6.168)</td>
<td>(3.345)</td>
</tr>
<tr>
<td>13th Year under Prohibition</td>
<td>1.454</td>
<td>2.663</td>
<td>-4.241</td>
<td>-6.195</td>
</tr>
<tr>
<td></td>
<td>(1.730)</td>
<td>(1.812)</td>
<td>(6.495)</td>
<td>(3.980)</td>
</tr>
<tr>
<td>14th Year under Prohibition</td>
<td>0.834</td>
<td>2.597</td>
<td>-1.574</td>
<td>-2.351</td>
</tr>
<tr>
<td></td>
<td>(1.377)</td>
<td>(1.679)</td>
<td>(7.353)</td>
<td>(3.880)</td>
</tr>
<tr>
<td>15th Year under Prohibition</td>
<td>-2.906</td>
<td>-0.381</td>
<td>-6.240</td>
<td>-2.561</td>
</tr>
<tr>
<td></td>
<td>(1.443)</td>
<td>(1.789)</td>
<td>(5.800)</td>
<td>(4.904)</td>
</tr>
<tr>
<td>16th Year under Prohibition</td>
<td>-3.146</td>
<td>-2.034</td>
<td>-2.467</td>
<td>-6.930</td>
</tr>
<tr>
<td></td>
<td>(1.675)</td>
<td>(1.730)</td>
<td>(6.333)</td>
<td>(2.853)</td>
</tr>
<tr>
<td>17th Year under Prohibition</td>
<td>-4.601</td>
<td>-2.282</td>
<td>3.591</td>
<td>-10.118</td>
</tr>
<tr>
<td></td>
<td>(1.605)</td>
<td>(1.655)</td>
<td>(6.486)</td>
<td>(3.316)</td>
</tr>
<tr>
<td>18th Year under Prohibition</td>
<td>-4.409</td>
<td>-3.033</td>
<td>1.493</td>
<td>-12.096</td>
</tr>
<tr>
<td></td>
<td>(1.691)</td>
<td>(1.218)</td>
<td>(6.955)</td>
<td>(2.976)</td>
</tr>
<tr>
<td>19th Year under Prohibition</td>
<td>-7.033</td>
<td>-8.162</td>
<td>2.012</td>
<td>-8.127</td>
</tr>
<tr>
<td></td>
<td>(2.290)</td>
<td>(2.682)</td>
<td>(7.097)</td>
<td>(3.293)</td>
</tr>
</tbody>
</table>

Time-varying Controls:
- Year Effects
- City Effects
- R squared
- No. of Observations

Notes: Constant included in all regressions is not reported. Standard errors are robust and clustered at the city level.
Time varying controls include log population, a Border indicator and a State-capital indicator.
Sample B is the balanced sample used for Structural estimation. Sample C includes all cities for which at least 8 years of data are available.

Effects of Heterogeneity in Moral Views during Prohibition

Table A4-2 now presents the estimated models from equation (3). Results are presented following the same structure as those in table A4-1, reporting the interaction terms only.
## Table A4-2: Moral Heterogeneity: Long-and Short Run Differences

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Homicide Rate per 100,000</th>
<th>Drunkenness Arrest Rate per 1,000</th>
<th>Police Expenditure Share</th>
<th>Per Capita Police Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Year under Prohibition x Wetness</td>
<td>-13.584 (14.735)</td>
<td>-11.161 (14.296)</td>
<td>-10.986 (21.368)</td>
<td>7.378 (8.716)</td>
</tr>
<tr>
<td>2nd Year under Prohibition x Wetness</td>
<td>-2.103 (11.105)</td>
<td>1.406 (9.480)</td>
<td>-13.652 (22.109)</td>
<td>12.240 (11.701)</td>
</tr>
<tr>
<td>3rd Year under Prohibition x Wetness</td>
<td>-5.829 (5.234)</td>
<td>-3.868 (8.489)</td>
<td>-5.762 (23.055)</td>
<td>19.051 (12.403)</td>
</tr>
<tr>
<td>4th Year under Prohibition x Wetness</td>
<td>9.867 (8.692)</td>
<td>8.950 (8.489)</td>
<td>3.567 (20.371)</td>
<td>23.345 (11.902)</td>
</tr>
<tr>
<td>5th Year under Prohibition x Wetness</td>
<td>8.545 (7.449)</td>
<td>11.051 (7.657)</td>
<td>-0.761 (20.323)</td>
<td>25.510 (11.646)</td>
</tr>
<tr>
<td>6th Year under Prohibition x Wetness</td>
<td>17.975 (9.142)</td>
<td>21.918 (9.429)</td>
<td>6.486 (20.533)</td>
<td>14.687 (12.512)</td>
</tr>
<tr>
<td>7th Year under Prohibition x Wetness</td>
<td>28.179 (11.930)</td>
<td>26.301 (7.865)</td>
<td>0.257 (21.481)</td>
<td>9.445 (11.501)</td>
</tr>
<tr>
<td>8th Year under Prohibition x Wetness</td>
<td>20.584 (7.737)</td>
<td>23.788 (7.566)</td>
<td>-7.061 (22.304)</td>
<td>3.815 (11.489)</td>
</tr>
<tr>
<td>9th Year under Prohibition x Wetness</td>
<td>32.922 (11.015)</td>
<td>32.455 (9.942)</td>
<td>-11.530 (22.904)</td>
<td>0.992 (11.363)</td>
</tr>
<tr>
<td>10th Year under Prohibition x Wetness</td>
<td>22.065 (9.146)</td>
<td>22.236 (8.214)</td>
<td>-3.044 (22.696)</td>
<td>1.578 (11.146)</td>
</tr>
<tr>
<td>11th Year under Prohibition x Wetness</td>
<td>29.034 (10.497)</td>
<td>29.363 (9.270)</td>
<td>-6.601 (23.652)</td>
<td>2.361 (13.084)</td>
</tr>
<tr>
<td>12th Year under Prohibition x Wetness</td>
<td>27.208 (7.488)</td>
<td>28.249 (6.853)</td>
<td>28.040 (28.151)</td>
<td>-1.229 (22.000)</td>
</tr>
<tr>
<td>13th Year under Prohibition x Wetness</td>
<td>17.276 (7.821)</td>
<td>15.742 (6.388)</td>
<td>18.806 (39.119)</td>
<td>-21.000 (28.344)</td>
</tr>
<tr>
<td>14th Year under Prohibition x Wetness</td>
<td>19.820 (6.940)</td>
<td>15.645 (6.432)</td>
<td>28.856 (83.417)</td>
<td>6.424 (43.458)</td>
</tr>
<tr>
<td>15th Year under Prohibition x Wetness</td>
<td>-1.986 (9.016)</td>
<td>0.901 (7.330)</td>
<td>-110.498 (74.038)</td>
<td>-214.430 (47.447)</td>
</tr>
<tr>
<td>16th Year under Prohibition x Wetness</td>
<td>17.982 (15.128)</td>
<td>6.842 (13.627)</td>
<td>-4.001 (66.963)</td>
<td>-46.025 (46.322)</td>
</tr>
<tr>
<td>17th Year under Prohibition x Wetness</td>
<td>-4.708 (8.743)</td>
<td>4.993 (7.385)</td>
<td>172.815 (155.191)</td>
<td>0.027 (0.019)</td>
</tr>
<tr>
<td>18th Year under Prohibition x Wetness</td>
<td>8.580 (7.886)</td>
<td>11.885 (7.348)</td>
<td>-22.205 (72.675)</td>
<td>0.011 (0.025)</td>
</tr>
<tr>
<td>19th Year under Prohibition x Wetness</td>
<td>7.646 (9.716)</td>
<td>14.108 (10.669)</td>
<td>-64.025 (50.044)</td>
<td>-0.014 (0.040)</td>
</tr>
</tbody>
</table>

### Notes:
- Constant included in all regressions is not reported. Standard errors are robust and clustered at the city level.
- Time varying controls include log population, a Border indicator and a State-capital indicator.
- Sample B is the balanced sample used for Structural estimation. Sample C includes all cities for which at least 8 years of data are available.
Prohibition Repeal

The repeal of the 18th Amendment itself also allows for the exploration of differential trends in criminality between cities with varying moral profiles. Here I exploit the repeal of nationwide Prohibition in December 1933 with the ratification of the 21st Amendment, to provide some additional evidence of the response of crime to Prohibition, and its stronger effects in communities with larger alcohol markets. I take advantage of the availability of more detailed crime data for the 1930-1936 period, taken from the Uniform Crime Reports (UCR) complied by the FBI starting in 1930. The UCR reports for a large number of cities, the total number of offences known to the authorities (which include any of the following: murder, rape, robbery, assault, burglary, larceny, and auto theft), and an independent measure of reported murders. Thus, I compare crime outcomes in the 1930-1933 period with the 1934-1936 period, allowing for differential behavior after repeal, as cities vary in their moral preference distribution, as proxied by \( \mu \). Indeed, simple summary statistics show that offences and murders were both lower in the post-18th Amendment years.

Thus, I look exclusively at the period 1930-1936, and run regressions for UCR offences and murders, and for the homicide rate.

\[
y_{ct} = \alpha_c + \beta_t + \delta CP_t + \phi C \cdot P_t \cdot \mu_c + \gamma' X_{ct} + \epsilon_{ct} \tag{33}
\]

where \( CP_t \) is an indicator variable for Constitutional Prohibition. Regression results are reported in table A4-3. Columns (1) – (4) look at the homicide rate. The coefficient on the interaction is always large, highly significant, and robust to the introduction of state-cross-year effects, suggesting that the fall in crime was larger in wetter cities. Take for example column (2). The estimates imply that for the city with mean “wetness” of 0.49, repeal was associated with a fall in the annual homicide rate of 4.6 = (0.49 \times 23.28) – 6.76. Even in the driest city, with \( \mu = 0.3 \), the estimated fall in the homicide rate is 0.21 = (0.3 \times 23.28) – 6.76. Columns (5) – (8) then present analogous results for the UCR number of murders per 100,000. The pattern is very similar to the one for the homicide rate, although standard errors increase somewhat, and the magnitude of the effect is smaller for the larger sample of cities covered. Nonetheless, for the sample for which homicide rates are available, results are very similar. The large standard errors for the sample in Columns (6) and (8) is due to the larger number of smaller cities included, in which reported murders were very small or close to zero, and present very little variation. Finally, Columns (9) – (12) present results for offences per 1,000. Interestingly, a pattern very similar to the one for homicides and arrests emerges, but this time, the effect is statistically significant especially in the larger sample including cities of smaller sizes. From Column (12), for example, it follows that repeal in the city with average “wetness” implied a fall in total offences of 3.85 = (0.49 \times 6.53) + 0.669 per 1,000 population, which is 43% of this variable’s standard deviation of 8.65. As the results suggest, while the reduction in criminality in larger cities was associated especially with a lower homicide rate, looking at a larger sample including smaller cities, repeal was associated with lower levels of other types of crime.

Table A4-3: Crime Fall after Repeal

---

46Average murders per 100,000 are 8.57 (s.e. = 10.3) in the 1930-1933 period, and 6.53 (s.e. = 8.8) in 1934-1936, with a t-statistic for the difference in means of 4.62. For offences per 1,000, the 1930-1933 mean is 16.26 (s.e. = 8.6), while the 1934-1936 mean is 15.6 (s.e. = 8.6), with a t-statistic of 1.64, significant at the 5% level.
Neighboring Markets

If individuals’ preferences are affected by the legal standard in place, say because they derive utility from abiding to the law, or, on the other hand, if individuals’ utility from taking an action increases when it is proscribed (a “forbidden fruit effect”), observed changes in the drunkenness arrest rate could be driven by these shocks in preferences. A way to isolate any taste shocks introduced by Prohibition is to look at the response of the alcohol market in a city which is already under Prohibition, when neighboring states’ prohibitionist status changes. If drinkers in a city under Prohibition have access to neighboring markets, which is very consistent with the concern of Prohibitionists of the time, and which motivated the passage of the Webb-Kenyon act, then the closure of neighboring markets should reduce the availability of liquor in the city, without having an effect on preferences. Thus, I collected information on the lengths of all state boundaries, and computed for each state, the share of state border in states under Prohibition at each point in time:

\[ SBP_{ct} = \frac{\sum_{j \in N_c} P_{jt} \times BorderLength_{cj}}{\sum_{j \in N_c} BorderLength_{cj}} \]

where \( P_{jt} \) is an indicator variable for state \( j \) being under Prohibition at time \( t \). \( N_c \) is the set of states neighboring city \( c \)’s state (e.g. \( N_{SanFrancisco} = \{Oregon, Nevada, Arizona, Mexico\}\)), and \( BorderLength_{cj} \) is the length in miles of the state boundary between city \( c \)’s state and state \( j \).

47 The importance of cross-state-boundaries alcohol trade after Prohibition was enacted in some states but not in neighboring ones is probably best exemplified by Daniel Okrent’s discussion of the huge traffic lanes along interstate 25, connecting Toledo, OH with Detroit, MI, after Michigan was covered by state-wide Prohibition in 1918. The highway was nicknamed “Avenue de Booze”. (See Okrent (2010, p. 107))

Isaac also highlights the importance of cross-state smuggling of alcohol after Tennessee started enforcing its Prohibition legislation: “The State bone-dry law, even when supplemented by the Reed amendment, or ‘national bone-dry law’, which made it a federal crime to transport intoxicants into a dry state, did not actually stop the flow of liquor into Tennessee. During 1917 and 1918, bootleggers where adequately supplied with whiskey brought from Kentucky to Nashville and Memphis by train, automobile, farm wagon, and river boat.” Isaac (1965, p. 254)

48 The information on state boundary lengths was taken from Holmes (1996). There are a total of 109 boundaries between U.S. states, and 16 international boundaries.

49 I include any international borders in the denominator, which amounts to considering Mexico and Canada as never being under Prohibition.
For the pre-Constitutional Prohibition period (1911-1919), when there is variation across states in Prohibition status, I estimate models of the form

\[ d_{ct} = \alpha_c + \beta_t + \delta P_{ct} + \eta SBP_{ct} + \phi P_{ct} SBP_{ct} + \gamma' X_{ct} + \epsilon_{ct} \] (34)

Table A4-4 presents the estimates of equation (34), for different specifications, and for samples A, B, and C. First, the fraction of border under Prohibition should have an effect on the drunkenness arrest rate only when the city itself is under Prohibition; otherwise the city's neighbors' Prohibition status should be irrelevant, since a free alcohol market is available. Thus, columns (1) – (3) in table A4-4 start presenting the estimates of a model where I include the share of border under Prohibition without an interaction with own Prohibition status. The share of state boundary under Prohibition is insignificant in the three specifications. Then columns (4) – (6) introduce the interaction term, and columns (7) – (9) additionally include time-varying controls (log of population, and time-varying state capital and South effects). The coefficient for the \( \phi \) is negative and large in magnitude, and always highly significant, except for column (8) when looking at the smaller B sample. The coefficient for \( \phi \) on column (6), for example, implies that a one pre-1920 standard deviation (0.29) increase in the fraction of state border under Prohibition implied a reduction in the drunkenness arrest rate of 1.93, which is 10% of the average pre-1920 drunkenness arrest rate in the sample. These estimates are very consistent with the idea that the sharp falls in drunkenness arrests observed were caused by a contraction in the alcohol supply available, and not due to preference shocks correlated with the introduction of Prohibition.

Table A4-4: Neighboring Prohibition

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Effect of Neighboring Prohibition</th>
<th>Drunkenness Arrests Rate per 1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A (1)</td>
<td>B (2)</td>
</tr>
<tr>
<td>Prohibition Indicator</td>
<td>-6.123 (1.376)</td>
<td>-6.072 (1.927)</td>
</tr>
<tr>
<td>Share of Border under Prohibition</td>
<td>-2.128 (2.138)</td>
<td>-6.257 (3.095)</td>
</tr>
<tr>
<td>Prohibition X Share of Border under Prohibition</td>
<td>-6.778 (2.650)</td>
<td>-1.348 (3.018)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time-varying Controls</th>
<th>No</th>
<th>No</th>
<th>No</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>City Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>K squared</td>
<td>0.226</td>
<td>0.319</td>
<td>0.228</td>
<td>0.234</td>
<td>0.320</td>
<td>0.226</td>
<td>0.248</td>
<td>0.353</td>
</tr>
<tr>
<td>No. of Cities</td>
<td>245</td>
<td>66</td>
<td>276</td>
<td>245</td>
<td>66</td>
<td>276</td>
<td>245</td>
<td>66</td>
</tr>
<tr>
<td>No. of Observations</td>
<td>1876</td>
<td>594</td>
<td>1861</td>
<td>1876</td>
<td>594</td>
<td>1861</td>
<td>1876</td>
<td>594</td>
</tr>
</tbody>
</table>

Notes: Constant included in all regressions is not reported. Standard errors are robust and clustered at the city level. Time varying controls include log population, a Border indicator and a State-capital indicator. Each specification is estimated for three samples: Sample A includes all city-x-year observations for which data is available. Sample B is the balanced sample used for Structural estimation. Sample C includes all cities for which at least 10 years of data are available. Coefficients for years under Prohibition and for the interactions between "wetness" and years under Prohibition not reported.

Dry Legislation

Can differences in pre-Prohibition alcohol-related legislation explain the trends in crime, arrests for drunkenness, and police enforcement? Prior to the adoption of state-level and nationwide Prohibition, different states had different types and numbers of dry laws. In fact, regulations over the alcohol market were in place almost everywhere. These included restrictions on selling hours, on the kinds of alcoholic beverages permitted, on the types of selling establishments allowed, and on taxation. There are two channels through which pre-Prohibition alcohol legislation might affect the
evolution of outcomes during Prohibition. First, given that early on during Prohibition collective
law enforcement decisions were likely to be closely related to initial “prior” beliefs about the policy’s
effects, variation in the short-run effects of Prohibition might be partly explained by variation in
pre-Prohibition dry legislation. The direction of an effect is not obvious a priori. On the one hand,
if these laws were being successful in shrinking the alcohol market and were not affecting crime,
people’s priors about the introduction of federal-level Prohibition could be very optimistic; on the
other hand, if the introduction of these laws was correlated with more crime, individuals might have
used this information to form negative priors about nationwide Prohibition. Second, differences in
dry laws could have created different initial conditions for the alcohol market at the time of Pro-
hibition adoption. For example, heavily regulated markets might have already developed a parallel
black market which could have eased the expansion of the illegal liquor trade during Prohibition.

To take a look at this question I reviewed the available information on state-level dry legislation
in the pre-18th Amendment period and constructed a variable counting the number of regulations
on the alcohol market at each point in time for each state. Interestingly, although the relationship
between average “wetness” of a state, as measured by \(\mu\), and the number of dry laws in place is
not very strong, it is actually positive. This is likely to be the result of the equilibrium political
strategies used by dry lobbies during the 1900s and 1910s. Because relatively “wet” regions were
unlikely to pass Prohibition laws, the lobbies focused their efforts on passing regulatory legislation
instead, which was politically feasible\(^5\). States like Michigan or Minnesota (both heavily “wet”),
passed, especially during the 1910s, significant amounts of regulatory legislation related to alcohol.
In the other extreme, radically “dry” states such as Utah and Oklahoma did not need to pass this
kind of legislation because they were already under Prohibition in the first place.

Pre-Prohibiton legislation is, of course, endogenous to outcomes over that period. Given that I want
to explore the effects of pre-Prohibition dry legislation on outcomes during Prohibition, which might
have an effect through initial beliefs (and hence, initial law enforcement choices during Prohibition),
or in how they shaped the local alcohol markets (and hence, in the subsequent response of alcohol
supply during Prohibition), below I briefly investigate the effect of pre-Prohibition legislation on
Prohibition outcomes, conditional on local preferences, by estimating models only for Prohibition
years, in which I allow for a differential effect of the number of pre-Prohibition dry laws over time
under Prohibition, controlling by a time-varying effect of baseline “wetness”:

\[
y_{ct} = \alpha_c + \beta_t + \sum_{\tau=1}^{k} \delta_{\tau} D_{ct} + \sum_{\tau=1}^{k} \eta_{\tau} D_{ct} L_{c} + \gamma' X_{ct} + \varepsilon_{ct} \tag{35}
\]

where \(L_{c}\) is the number of dry laws in place right before the city is under Prohibition, and \(X_{ct}\)
includes interactions of \(\mu_{c0}\) with year indicators. Because these models only look at years under
Prohibition, I omit the indicator for \(\tau = 1\), so the interpretation of the “years under Prohibition”
indicator variables is different; coefficients must now be interpreted as relative to having experienced
Prohibition for one year. The \(\eta_{\tau}\)'s should capture any time-varying differential effects of an extra
piece of dry pre-Prohibition legislation on Prohibition outcomes. Flexibly controlling for the moral

\(^5\) The data on dry legislation was mostly taken from Cherrington (1920) and League (1932). Both sources have a
detailed and comprehensive compilation of dry legislation during these decades.
profile of the city as proxied by \( \mu \) is important given that pre-Prohibition dry legislation is likely to be correlated with preferences in the city. To save space, in table A4-5 I only present results for the coefficient estimates for the \( \eta_i \)'s of the benchmark fixed effects specifications. Regression results do not show any significant relationship between the amount of pre-Prohibition dry legislation and the homicide rate or the arrest rate at any time during Prohibition. There also appears to be no relation between these laws and the behavior of per capita expenditure in policing during Prohibition years. For the expenditure share, on the other hand, the interaction terms are small in magnitude to be correlated with preferences in the city. To save space, in table A4-5

Table A4-5: Effects of pre-Prohibition Dry Legislation

<table>
<thead>
<tr>
<th>Year Effects</th>
<th>Controls</th>
<th>Dependent variable</th>
<th>Dry Laws</th>
<th>Per Capita Police Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Homicide Rate per 100,000</td>
<td>Drunkenness Arrests Rate per 1,000</td>
<td>Police Expenditure Share</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
</tr>
<tr>
<td>2nd Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.208 (0.020)</td>
<td>-0.234 (0.291)</td>
<td>-0.199 (0.005)</td>
<td>0.002 (0.003)</td>
</tr>
<tr>
<td>3rd Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.247 (0.019)</td>
<td>-0.237 (0.205)</td>
<td>0.120 (0.006)</td>
<td>0.001 (0.002)</td>
</tr>
<tr>
<td>4th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.218 (0.029)</td>
<td>-0.236 (0.282)</td>
<td>0.163 (0.009)</td>
<td>0.001 (0.003)</td>
</tr>
<tr>
<td>5th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.229 (0.028)</td>
<td>-0.427 (0.280)</td>
<td>0.122 (0.009)</td>
<td>0.000 (0.004)</td>
</tr>
<tr>
<td>6th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.229 (0.028)</td>
<td>-0.509 (0.250)</td>
<td>0.139 (0.005)</td>
<td>0.008 (0.003)</td>
</tr>
<tr>
<td>7th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.229 (0.028)</td>
<td>-0.431 (0.250)</td>
<td>0.129 (0.005)</td>
<td>0.004 (0.004)</td>
</tr>
<tr>
<td>8th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.468 (0.200)</td>
<td>-0.010 (0.009)</td>
<td>0.001 (0.007)</td>
</tr>
<tr>
<td>9th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.672 (0.200)</td>
<td>-0.054 (0.009)</td>
<td>0.002 (0.000)</td>
</tr>
<tr>
<td>10th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>11th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>12th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>13th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>14th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>15th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>16th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>17th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>18th Year under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>19th and more Years under Prohibition x Pre Prohibition Dry Laws</td>
<td>0.275 (0.034)</td>
<td>-0.415 (0.200)</td>
<td>-0.056 (0.009)</td>
<td>0.000 (0.000)</td>
</tr>
</tbody>
</table>

Notes: Constant included in all regressions is not reported. Standard errors are robust and clustered at the city level. Time-varying controls include big population, a border indicator and a State-capital indicator. Sample B is the balanced sample used for structural estimation. Sample C includes all cities for which at least 8 years of data are available. Coefficients for years under Prohibition and for the interactions between "wetness" and years under Prohibition not reported.
Women’s suffrage

Several historians have attributed some of the success of Prohibition in the United States to the significant role that the Women’s Suffrage Movement played. It is undeniable that women played a prominent role in the conflict over alcohol consumption, and were of importance at least since the 1870s, when a group of Ohio women began the “Temperance Crusade” that spread throughout all of the Midwest. A group of women would visit the area’s saloons one by one, and protest and pray for days until the owners decided to close. The long-term effects of the crusade are likely to have been minimal, but it was the first major women-specific social mobilization, and was the origin of the WCTU some years later. In the Twentieth century, both the Women’s Suffrage Movement and the Temperance Movement were part of the Progressive-era reforms, and organizations such as the WCTU were involved in the political struggle around both issues. Although U.S.-wide women’s suffrage (19th Amendment) was ratified into the Constitution in 1920, after the adoption of nationwide Prohibition (18th Amendment), authors such as Okrent (2010) argue that the Women’s Suffrage Movement gave a major impulse to the Prohibition movement. The almost simultaneous ratification of the 18th and 19th Amendments makes it impossible to identify any specific effects that women’s suffrage might have had during federal Prohibition years. Nonetheless, prior to the 19th Amendment several states had already extended the franchise to women. As a way to explore the importance of women’s enfranchisement on Prohibition-related outcomes, I exploit the variation in women’s suffrage enfranchisement prior to 1920, when both the 18th and 19th Amendments were ratified, to see if Prohibition had differential effects in cities with and without women’s suffrage. If the distribution of women’s preferences over Prohibition enforcement was different than men’s, cities allowing women’s suffrage could be under a differential trend. Thus, for the 1910-1919 period, I run regressions of the form

\[ y_{ct} = \alpha_c + \beta_t + \eta W_{ct} + \sum_{\tau=1}^{k} \delta_{\tau} D_{\tau} + \sum_{\tau=1}^{k} \phi_{\tau} D_{\tau} W_{ct} + \gamma' X_{ct} + \varepsilon_{ct} \] (36)

In equation (36), \( W_{ct} \) is an indicator variable taking the value of 1 if city \( c \) has women’s suffrage in year \( t \). Table A4-6 presents results of the estimates of the \( \phi_{\tau} \)’s from equation (36) for the different outcome variables, in the specifications including city fixed effects, time-varying controls, and year effects. The regressions include only up to \( \phi_5 \), because before 1919 no city with women’s suffrage in the sample had experienced more than 5 years under Prohibition. There is no evidence of a differential trend in the homicide rate in cities with women’s suffrage. This is unsurprising given that the short-run effects of Prohibition on the homicide rate were very small. For the outcomes which did have large short-run changes after the introduction of Prohibition, if anything, Columns (2) – (3) in table A4-6 show that the introduction of women’s suffrage is correlated with more drunkenness arrests in the short run (after two to three years under Prohibiton), but the net effect is small and insignificant quickly thereafter. This result is also not robust to the restricted \( B \) sample (column (2)). When looking at police enforcement in Columns (4) – (7), the results are also very inconclusive.

---

51 Women’s Suffrage prior to the 19th Amendment was adopted by the states as follows: Wyoming in 1869, Colorado in 1893, Utah and Idaho in 1896, Washington in 1910, California in 1911, Arizona, Kansas and Oregon in 1912, Montana and Nevada in 1914, New York in 1917, and Michigan, Oklahoma and South Dakota in 1918.
During years with women’s suffrage, cities have slightly lower but insignificant policing, which is actually inconsistent with the idea that women’s anti-Prohibitionism should translate to higher law enforcement and a smaller alcohol market after their enfranchisement. Overall, the available evidence does not suggest that alternative legislation, such as dry laws or women’s suffrage, might have been driving the trends in law enforcement, crime and arrests presented in Section 4.1.

Table A4-6: Effect of Women’s Suffrage

<table>
<thead>
<tr>
<th></th>
<th>Women's Suffrage Indicator</th>
<th>Homicide Rate per 100,000</th>
<th>Drunkenness Arrests Rate per 1,000</th>
<th>Per Capita Police Expenditure</th>
<th>Police Expenditure Share</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Women's Suffrage</td>
<td>0.1171</td>
<td>-1.080</td>
<td>-2.183</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(1.024)</td>
<td>(1.696)</td>
<td>(1.132)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>1st Year under Prohibition x Women's Suffrage</td>
<td>-0.7345</td>
<td>1.410</td>
<td>1.311</td>
<td>-0.007</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(1.502)</td>
<td>(2.812)</td>
<td>(1.623)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>2nd Year under Prohibition x Women's Suffrage</td>
<td>-2.3699</td>
<td>-5.454</td>
<td>5.409</td>
<td>-0.017</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(4.357)</td>
<td>(3.447)</td>
<td>(2.645)</td>
<td>(0.006)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>3rd Year under Prohibition x Women's Suffrage</td>
<td>-1.3185</td>
<td>-4.172</td>
<td>5.616</td>
<td>-0.021</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(4.537)</td>
<td>(3.969)</td>
<td>(3.655)</td>
<td>(0.009)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>4th Year under Prohibition x Women's Suffrage</td>
<td>5.5515</td>
<td>-0.017</td>
<td>4.276</td>
<td>-0.036</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(9.094)</td>
<td>(6.215)</td>
<td>(3.767)</td>
<td>(0.010)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>5th Year under Prohibition x Women's Suffrage</td>
<td>3.329</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Time-varying Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| City Effects          | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| R squared             | 0.108 | 0.344 | 0.291 | 0.310 | 0.195 | 0.649 | 0.606 |
| No. of Cities         | 66  | 66  | 236  | 66  | 217  | 66  | 217  |
| No. of Observations   | 564 | 554 | 1861 | 519 | 1427 | 528 | 1427 |

Notes: Constant included in all regressions is not reported. Years Under Prohibition Indicators not reported. Standard errors are robust and clustered at the city level. Time varying controls include log population, a Border indicator and a State-capital indicator. Sample B is the balanced sample used for Structural estimation. Sample C includes all cities for which at least 10 years of data are available.

Selection in the Public Opinion data

A caveat in the elections data is that several states including Louisiana, New Jersey, New York, and Pennsylvania, did not hold any liquor-related referendum in a pre-Prohibition year. This induces a potential selection bias in the estimates of equation (4) because these states never held a referendum regarding liquor precisely due to the highly anti-Prohibitionist preferences of their citizens. As a robustness check I also estimate a selection model, by specifying a selection equation for holding a referendum (at the state level). More specifically, I assume that

\[ r_{St} = \begin{cases} 
1 & \text{if } t = 0 \text{ and } \eta' Z_{S0} + v_{S0} > 0 \\
1 & \text{if } t = 1 
\end{cases} \]

where \( r_{St} \) is an indicator variable for state \( S \) holding a liquor referendum, \( Z_{S0} \) includes the state’s share of adherants to a wet religion and the share of native white individuals in 1910, and \( v_{S0} \sim N(0,1) \), with \( E[\varepsilon_{ct}|v_{S0}] = \mu v_{S0} \) and \( E[\varepsilon_{ct}|v_{S0}] = 0 \). This implies that

\[ E[w_{ct}|\mu_{ct}, \mu_{0t}, X_{ct}, r_{St} = 1] = \alpha_c + \beta t + \delta \mu_{ct} + \phi \mu_{0t} + \gamma' X_{ct} + \kappa \lambda (\eta' Z_{S0}) 1_{t=0} \]

where \( \lambda() \) is the inverse Mills ratio. Results are reported in columns (6), (12), and (18) of table 3.
Appendix 5: Data Sources

Most of the information available for the study of Prohibition is available at the city level, so I focused on constructing a yearly panel dataset of cities, covering the 1910s, '20s and early '30s. The data collected comes from a wide array of sources. The first source of information is the collection of original documents from the National Commission on Law Observance and Enforcement, most commonly known as the Wickersham Commission after the name of its Chair Commissioner, Attorney General George Wickersham. It was appointed in the Spring of 1929 by President Hoover, with the specific purpose of "studying exhaustively the entire problem of the enforcement of our laws and the improvement of our judicial system, including the special problem and abuses growing out of the Prohibition laws" (Wickersham-Commission (1928-1931b)). It was, of course, appointed as a response to the growing concerns about the effects Prohibition was having throughout the country, and the public discontent over the policy’s effects. The Commission produced a series of reports on the different aspects of Prohibition, after directly collecting data and evidence across the country, and issued its main findings in 1931. Harvard’s Law School Library currently holds the collection of documents from the Commission, including the originals of much of the summarized data in the published reports, in addition to several other unpublished information. The detailed city-by-city “Prohibition Survey” reports, compiled directly by commissioners traveling to the cities and collecting information about the recent evolution of criminality, and the “Cost of Crime” state-level folders, providing detailed data on local law enforcement activity, contain the most valuable information from the Wickersham papers.

Law Enforcement Data

Other than the data mentioned in section 3, the Wickersham Commission papers also contain other data on total arrests, unfortunately, available only during the 1910s and in 1929. Data on a set of other Prohibition enforcement outcomes is available only at the state level from the U.S. Bureau of Prohibition for the years 1923-1932, such as the number of still and liquor seizures, arrests of alcohol producers, and casualties caused by Prohibition enforcement agencies (see table 1). This information aggregates Prohibition enforcement operations from both federal and local authorities in most cases. I collected data on criminal judicial prosecutions from the Attorney General Annual reports, which are available at the Judicial District level only, for the years 1915-1936.

Demographic and Religious Data

For the first four decades of the Twentieth century, data on the distribution of religious ascriptions is available from the 1906, 1916, 1926, and 1936 decennial Censuses of Religions. The Censuses have comprehensive information about the number of adherents to each of the different faiths or churches in the United States.
Public Opinion Data

Most of the data comes from the state official rosters or “blue books”, which states publish on an annual or biannual basis. The information for some of the states was found in the state archives, and for a few other referenda not reported in official sources, I took the data from local newspapers. A second major source of electoral data on the Prohibition issue are the election returns for the 21st Amendment Constitutional Convention elections, also found in the state rosters and some state archives.

Structural Estimation Data

The sample includes cities from all over the United States, and although the range of population sizes in this sample of cities goes from 51,000 to 5.6 million (1920 numbers), admittedly this is a sample of urban communities. Of course, this is mainly due to the availability for the homicide rate data, which was reported on a population basis and for cities only. It is important to stress that the results should be seen as the effects of Prohibition in the most urbanized parts of the American society.

Table A5-1: Sample of Cities in the Structural Estimation

<table>
<thead>
<tr>
<th>city</th>
<th>state</th>
<th>city</th>
<th>state</th>
<th>city</th>
<th>state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akron</td>
<td>OH</td>
<td>Indianapolis</td>
<td>IN</td>
<td>Portland</td>
<td>OR</td>
</tr>
<tr>
<td>Albany</td>
<td>NY</td>
<td>Jersey City</td>
<td>NJ</td>
<td>Providence</td>
<td>RI</td>
</tr>
<tr>
<td>Atlanta</td>
<td>GA</td>
<td>Kansas City</td>
<td>KS</td>
<td>Reading</td>
<td>PA</td>
</tr>
<tr>
<td>Baltimore</td>
<td>MD</td>
<td>Kansas City</td>
<td>MO</td>
<td>Richmond</td>
<td>VA</td>
</tr>
<tr>
<td>Birmingham</td>
<td>AL</td>
<td>Los Angeles</td>
<td>CA</td>
<td>Rochester</td>
<td>NY</td>
</tr>
<tr>
<td>Boston</td>
<td>MA</td>
<td>Louisville</td>
<td>KY</td>
<td>Saint Louis</td>
<td>MO</td>
</tr>
<tr>
<td>Bridgeport</td>
<td>CT</td>
<td>Lowell</td>
<td>MA</td>
<td>Saint Paul</td>
<td>MN</td>
</tr>
<tr>
<td>Buffalo</td>
<td>NY</td>
<td>Memphis</td>
<td>TN</td>
<td>Salt Lake City</td>
<td>UT</td>
</tr>
<tr>
<td>Cambridge</td>
<td>MA</td>
<td>Milwaukee</td>
<td>WI</td>
<td>San Antonio</td>
<td>TX</td>
</tr>
<tr>
<td>Camden</td>
<td>NJ</td>
<td>Minneapolis</td>
<td>MN</td>
<td>San Francisco</td>
<td>CA</td>
</tr>
<tr>
<td>Chicago</td>
<td>IL</td>
<td>Nashville</td>
<td>TN</td>
<td>Scranton</td>
<td>PA</td>
</tr>
<tr>
<td>Cincinnati</td>
<td>OH</td>
<td>New Bedford</td>
<td>MA</td>
<td>Seattle</td>
<td>WA</td>
</tr>
<tr>
<td>Cleveland</td>
<td>OH</td>
<td>New Haven</td>
<td>CT</td>
<td>Spokane</td>
<td>WA</td>
</tr>
<tr>
<td>Columbus</td>
<td>OH</td>
<td>New Orleans</td>
<td>LA</td>
<td>Springfield</td>
<td>MA</td>
</tr>
<tr>
<td>Dallas</td>
<td>TX</td>
<td>New York</td>
<td>NY</td>
<td>Syracuse</td>
<td>NY</td>
</tr>
<tr>
<td>Dayton</td>
<td>OH</td>
<td>Newark</td>
<td>NJ</td>
<td>Toledo</td>
<td>OH</td>
</tr>
<tr>
<td>Denver</td>
<td>CO</td>
<td>Norfolk</td>
<td>VA</td>
<td>Trenton</td>
<td>NJ</td>
</tr>
<tr>
<td>Detroit</td>
<td>MI</td>
<td>Oakland</td>
<td>CA</td>
<td>Washington</td>
<td>DC</td>
</tr>
<tr>
<td>Fall River</td>
<td>MA</td>
<td>Omaha</td>
<td>NE</td>
<td>Wilmington</td>
<td>DE</td>
</tr>
<tr>
<td>Grand Rapids</td>
<td>MI</td>
<td>Paterson</td>
<td>NJ</td>
<td>Worcester</td>
<td>MA</td>
</tr>
<tr>
<td>Hartford</td>
<td>CT</td>
<td>Philadelphia</td>
<td>PA</td>
<td>Yonkers</td>
<td>NY</td>
</tr>
<tr>
<td>Houston</td>
<td>TX</td>
<td>Pittsburgh</td>
<td>PA</td>
<td>Youngstown</td>
<td>OH</td>
</tr>
</tbody>
</table>

Section 6.1 mentioned that in spite of being a dynamic model, Maximum Likelihood estimation was not subject to an initial conditions problem. The careful reader might have noticed that this requires the sample to cover years under no Prohibition and under Prohibition, while a few states were already
under Prohibition before 1911. Given the timing of the adoption of Prohibition across States (see figure 3), and the data availability, for Nashville and Memphis in Tennessee, Atlanta in Georgia, and Kansas City in Kansas, the sample covers Prohibition years exclusively. These three states officially adopted Prohibition in 1909, 1908 and 1880, respectively. Nevertheless, following the historical account on Prohibition in Tennessee, I code the cities in this state as being under Prohibition only starting in 1914. As mentioned in footnote 5.6.5, the governor of Tennessee decided not to enforce the constitutional amendment enacted in 1909, and Prohibiton only was enforced after the new Republican governor took office.5

Although for Altanta, GA, and Kansas City, KS, the drunkenness arrests data also shows a fall only in 1917 (when War-time prohibition was adopted), suggesting little actual law enforcement of the state Prohibition laws (Atlanta’s arrests fall from 18.4 to 12.2 between 1916 and 1917), there is no clear evidence that the laws were actually not being enforced. Instead of specifying a distribution for the unobserved homicide rate prior to 1911 for these three cities, I estimate the structural model assuming they enter Prohibition in 1917, and check the robustness of the model to excluding them from the estimation altogether.

---

5The fact that Prohibition did not take place in Tennessee before 1914 can be corroborated directly by looking at the drunkenness arrests data. For example, this variable falls from 17.5 to 8.9 per 1,000 people between 1913 and 1914 in Knoxville. Hilary House, Nashville’s mayor at the time, even explicitly “acknowledged before the world that the state-wide Prohibition law is violated in Nashville... with knowledge and consent of the great majority of the people”. Isaac (1965, p. 174)
Chapter 2: Entrenchment (and Encroachment) Dynamics*

Camilo Garcia-Jimeno†

May, 2011

Abstract

This paper studies the strategic interaction between competition and ratchet effect incentives in the context of a dynamic coalition formation game of incomplete information. When the support decisions of potential coalition members can reveal private information about the cost of providing political support, reluctance to accept informative offers naturally arises. On the other hand, a ruler (the player making coalition proposals) can try to exploit the competitive nature of the coalition formation game to give revelation incentives to potential coalition members. Nevertheless, offers that are informative must imply risk of triggering a political transition. The interplay between these forces determines the extent to which the ruler can extract information, and his ability to entrench (reduce the likelihood of a political transition) and encroach (increase his share of rents) over time. By restricting attention to the Markov Perfect Bayesian equilibria of this game, I find that the ruler can exploit competition to extract some information, but that ratchet effect incentives do limit the extent to which he can do it. I show that the overall dynamics are strongly shaped by an informational commitment problem that arises when a ruler’s beliefs are “too optimistic”. A ruler that is tempted to make risky offers in the future will find it harder to extract information in the present because the likelihood of a future political transition increases the value of denying support for potential coalition members. Counterintuitively, this implies that power transitions are likely to happen when public beliefs are relatively optimistic. Moreover, the amount of information that can be extracted does not vary monotonically with the degree of optimism of public beliefs.

---

*This paper constitutes the second chapter of my Ph.D. dissertation at MIT. I am very grateful to Daron Acemoglu and James Robinson for their advice and guidance.

†MIT, Department of Economics. cgarcia@mit.edu. 50 Memorial Drive, Of. E52-201, Cambridge, MA 02139. This paper is the second chapter of my PhD Dissertation at the Massachusetts Institute of Technology. I am extremely grateful to my advisors Daron Acemoglu, James Robinson and James Snyder for their guidance and support.
1 Introduction

In any political environment, from weakly to highly institutionalized, individuals or groups in power are constantly engaged in the business of maintaining the political support that allows them to remain in power. At the same time, holding office becomes a source of benefits for those who hold it. A natural tension arises between remaining in power and perceiving its benefits, since gathering political support requires distributing some of the benefits from holding office among heterogeneous groups. On the other hand, politically powerful groups who trade benefits for political support face the opportunity cost of forgoing the possibility of becoming the ones in power. As a result, such an environment naturally gives incentives for groups with political power to be reluctant in supporting incumbent groups, and suggests that we should observe either a high degree of churning or a very equitable distribution of benefits across the politically powerful groups that constitute a given polity.

In sharp contrast, many historic and contemporary societies are characterized by quite the opposite: specific groups or individuals remaining in power for extended periods of time (this I will call a high degree of entrenchment), and a highly unequal distribution of benefits across groups (this I will call a high degree of encroachment). The nature of this problem is not exclusively a feature of societies with a low degree of political development; in fact, one possible way of rationalizing formal officeholding tenure limits frequently observed in more developed polities, is as an explicit way to address the problem of political entrenchment. Particularly puzzling is the fact that those groups granting political support to an incumbent could potentially achieve office for themselves by denying it to the current ruler. This is especially striking in light of the fact that entrenchment and encroachment often evolve in tandem. Some regimes manage to increasingly encroach over time (for example, through the undermining of the separation of powers, the centralization of decision-making, or even through outright theft) under the consent of their political support bases, and sustained in power precisely by those same groups who would benefit from replacing them. If an incumbent is becoming increasingly encroached, why do politically powerful non-incumbent groups not oust him? How is it that potential replacements do not take their political support away from the ruler as they see their own gains from giving such support fall over time? What strategies do groups in power use to achieve this result? What are the limits on these dynamics? When are political transitions more likely to take place?

Evidence of the lack of a sound theoretical understanding of this problem is Myerson’s claim that “... a leader who was expected to subsequently rule as an arbitrary tyrant would be unable to recruit supporters for his original rise to power... our analysis offers a reason to view such tyrannies as exceptional cases that necessarily involve a failure of rational expectations by early supporters of the regime” (Myerson (2008, p. 136)). Nevertheless,
the political economy literature has offered several rational choice-based explanations for
the stability of undelivering political regimes, to which this paper relates. In earlier work
discussing why “empirically the Junta characteristically shrinks to one man”, Tullock (1987)
proposes a process of rounds in which the victorious politician at each round accumulates
more power than his fellows and is then able to exclude them from the ruling coalition in
future rounds. The process then leads to one-person rule and a high degree of concentration
of political power. On the other hand, Egorov and Sonin (2005) argue that dictators choose
incompetent subordinates to reduce the probability of betrayal, because more competent
viziers have more incentives to replace the ruler. Dictators, thus, must trade-off loyalty for
competence to remain in power. In Svolik (2008), a leader can take hidden actions to weaken
his ruling coalition, which can respond by staging a costly coup. The unobservability of
the action together with the costliness of setting up a coup create the possibility that in
equilibrium, the ruler succesfully gets entrenched (becomes an autocrat).

The theories above have one element in common: entrenchment requires rulers to success-
fully alter the balance of political power by weakening their opponents. This is a natural
first approach to think about entrenchment, but at some level it begs the more difficult
question of how an incumbent might be able to entrench and encroach in spite of other
groups holding enough political power to oust him. Some recent work has tackled this ques-
tion. Padro-Miquel (2007) argues that the paradox of strong and stable autocratic regimes
in African ethnically divided societies can be explained by a strategy of “politics of fear”.
In a society consisting of two groups, one of which holds power, coethnics of the current
ruler fear doing much worse under the competing ethnic group, and as a result are willing
to support him despite his predatory behavior. In a similar vein, Acemoglu et al. (2004)
suggest that the stability of kleptocratic regimes might be explained by the ability of rulers
to undermine the collective action of groups who could potentially contest them through
credible punishments to opponents and prizes to supporters. Finally, in their “selectorate”
model Bueno-De-Mezquita et al. (2005) emnphazise the importance of the size of the winning
coalition relative to the set of potential coalition members to understand the incentives for
supporting rulers and the strategies that rulers might adotp. These models do not incor-
porate very rich dynamics, in part because of Myerson’s concern above. In a setting where
the source of political power of competing groups is independent from the ruler’s actions, a
history in which political opponents initially grant their support to a ruler and subsequently
get expropriated, while their chances to gain office shrink, cannot be subgame perfect; the
ruler would never achieve the political support to remain in power in the first place. As a
result, these models do not feature trajectories of increasing entrenchment or encroachment
over time. Since the beginning of the game, incumbents are fully entrenched and encroached.

There is, nevertheless, some recent literature in dynamic legislative bargaining in which an
increasingly uneven distribution of resources can obtain as bargaining goes on over time (see in particular Kalandrakis (2004) and Kalandrakis (2009)). This result relies on the restrictive assumption of a state-dependent process for the status quo distribution. Moreover, these models do not allow for the possibility of political transitions, so that players in those games are effectively unable to react to unfavorable offers.

This paper suggests an alternative view by proposing a specific theoretical framework to analyze the dynamics of political entrenchment and encroachment, which intends to fill in the previously highlighted gaps in the literature. The analysis is rooted on several ingredients which I believe are key for the understanding of this set of issues. First, I argue that it is important to make an explicit distinction between political power as an asset (the capacity of political players to threaten the stability of the ruler’s incumbency, for example due to charisma or the ability to deliver patronage or exercise violence), and the resulting distribution of benefits from the political game. Benefits and political power might be highly correlated, but are conceptually distinct. To make such distinction explicit, I will hold fixed the distribution of political power and will study how it might map into the distribution of benefits, across groups and over time, while abstracting from the effect that the distribution of benefits might have on the subsequent distribution of political power. Using the previous language, this paper will look at how entrenchment leads to encroachment, but not, the other way around, precisely to study a scenario where the ability of competing groups to oust an incumbent is not falling over time. The premise is that there is no such thing as an almighty ruler.

Second, that any political coalition formation game is in essence, a bargaining game. Following the huge literature on bargaining and reputations (see for example Fudenberg et al. (1985), Hart and Tirole (1988), Sobel and Takahashi (1983), Bar-Isaac (2003), Schmidt (1993)), I argue that a key interaction between incumbents and potential supporters in a coalition formation game is the constant effort of incumbents to learn how cheaply they can buy the political support needed to remain in power. Because a coalition formation game is, by its very nature, a non-anonymous interaction, potential supporters, on the other hand, are constantly trying to look expensive. Thus, learning might be important in the evolution of political interactions within coalition formation games.

This, in fact, is based on a third tenet of this paper: that providing political support is costly. Depending on the sources of political power in a specific context, this could be, for example, because leaders have to mobilize armies, or to pander their constituencies through clientelistic exchanges. But while the amount of political power held by different groups in society is easily observable, for example because it is closely related to population size or other resources, the cost of providing political support is hard to observe for outsiders, since it is likely to vary across groups and depend on their local characteristics. Some evidence of
across-groups variation in the cost of providing political support comes from McMillan and Zoido (2004). They document the episode of rampant corruption in Peru during the Fujimori administration in the 1990s, when the President’s main advisor and Chief of Intelligence systematically engaged in the trade of money for political support to the regime. The data reveals that support from three groups was highly valued by the government: the courts, the military, and the media. Interestingly, the bribes paid to media owners were orders of magnitude higher than those paid to judges or generals. Of course, it is likely that these groups varied in the amount of political power they could sell to the Fujimori regime, but the authors highlight that media outlets faced high reputational costs of engaging in the corrupt deals, and hence, required larger bribes.

The main mechanism I investigate in this paper is the interplay between political competition within a coalition formation game, and the ratchet effect incentives that arise as a result of the possible exploitation of valuable information revealed over time. I address the question of the extent to which an incumbent ruler can exploit the competitive nature of the coalition formation game to undermine the ratchet effect incentives of potential coalition members. Learning about a given group’s cost of providing political support has two direct benefits for the ruler. If the information revealed is that the group has a low cost, this will allow him to make lower offers in the future, and hence, increase the extent of his encroachment. Moreover, by increasing the precision of his beliefs about this cost, he will be able to make offers that imply a lower risk of being rejected, an hence increase the extent of his entrenchment. Of course, any given group should be reluctant to accept a low offer, both because it is unattractive relative to the possibility of ousting the ruler and triggering a political transition, but moreover, because low offers are likely to be informative, leading to low offers in the future and hence, implying a ratchet effect.

When unanimity is not necessary for the ruler to remain in power, an incumbent ruler might try to exploit competition within the coalition formation game to play potential coalition members against each other, giving them incentives to reveal information. Competition might be able to undermine the ratchet effect, and if this is so, a ruler will be able to entrench and encroach over time, as information becomes more precise. The model will be informative about the conditions under which exploiting competition can be used as an information extraction mechanism, and the endogenous limits that arise on the use and effectiveness of this strategy. Notice that a group is willing to deny support to the incumbent if it expects the competing group to do so too. Otherwise, a political transition will not ensue, while the transfer will have been forgone. As a result, beliefs are of strategic importance not only for the ruler, but also for other competing groups. In the absence of competition, a given group should prefer the ruler to be very pessimistic about his cost. Under competition, if the competing group is more likely to be included in a coalition when the ruler is more
optimistic about him, the group might actually prefer to reveal his private information. Thus, as information is acquired it becomes more risky to play a non-supporting strategy. At the same time, if the ruler wants to increase his probability of remaining in power, he will have to make larger offers which, by the fact of being higher, are also less informative. The ruler has to trade-off rents for risk and learning. The possibility of learning in this context introduces interesting entrenchment and encroachment dynamics, absent in the models discussed above, where outcomes are stationary.

The importance of the ratchet effect as a source of inefficiency has been highlighted in the industrial organization literature, and its main results are well known (see for example Freixas et al. (1985) and Laffont and Tirole (1988), or Choi and Thum (2003) for a more recent application to corruption). Moreover, the strategic interaction between competition and the ratchet effect also has been studied. In recent research on the subject, Charness et al. (2010) argue that workers face ratchet effect incentives on the job, when managers raise performance standards as they learn about worker productivity, leading to underprovision of effort in the first place. They argue that competition between workers in the labor market can eliminate the ratchet effect, and show experimental results consistent with this idea. To my knowledge, the theoretical relationship between competition and the ratchet effect has not been explored in the political economy literature.

The model I propose is intended to capture this strategic environment in the simplest and more “institutions-free” possible way. Each period, a ruler faces two groups, each with private information about their cost of providing political support, and needs the support of at least one of them to remain in power until the next period. The ruler makes public sequential offers, which are then accepted or rejected, and beliefs about the cost of political support are updated rationally as a function of the informational content of the ruler’s offers and the support decisions of the potential coalition members. If the ruler does not achieve the support of at least one group, he is ousted and a political transition takes place. The fact that beliefs about the groups’ cost types evolve over time as a function of players’ actions makes this not a repeated, but rather a dynamic game, where the state variable is the vector of public beliefs. Equilibrium will require that posterior beliefs are consistent with players’ strategies.

The model delivers several predictions and suggests some non-trivial intuitions. In the first place, in opposition to the regulation literature where the ratchet effect appears as a source of inefficiency, in a political economy context it might actually be desirable, to the extent that it is capable of limiting entrenchment and encroachment of incumbents. The overall equilibrium dynamics of the model are strongly shaped by an informational commitment problem faced by the ruler, which has both static and dynamic implications on his ability to extract information. A naive intuition would suggest that a ruler with more optimistic
beliefs (a high probability that a given potential coalition member is low cost) should be able to make more precise offers, and thus, be entrenched more easily. In contrast, when beliefs are very optimistic, the ruler is willing to make highly-informative offers and incur the risk of being ousted, because the likelihood of a rejection will be low. But the sequential nature of the offers implies that whomever is made a first offer, knowing the ruler will be tempted to extract information from the competing group, will find the value of rejecting support to be higher. As a result, the ruler will only be able to get the support of this group by making a high and uninformative offer, which reduces encroachment. This is a commitment problem for the ruler, because he would actually benefit if he could commit not to learn about the second group’s cost after a rejection by the first group. On the other hand, this commitment problem has different static and dynamic implications. Although it implies that a ruler will either have to forgo learning or risk being ousted, effectively leaving rents for the competing groups, it also implies that the ruler will be able to extract information very effectively if in the future he will face such a commitment problem, as the competing groups will be very motivated to accept offers that subsequently make beliefs so optimistic that the ruler will face the commitment problem.

For pessimistic enough beliefs, on the other hand, naive intuition would suggest little entrenchment, since the probability of rejection of an informative offer is large. Nevertheless, in such case the ruler is unwilling to incur such high risk, and prefers to remain in power for sure at the expense of information extraction. As a result, incumbents face no commitment problem in this case, since the first group to be offered knows that following a rejection, the ruler will credibly make a high enough offer to the competing group as to remain in power. This will allow the ruler to exploit competition to extract a limited amount of information. Pessimistic rulers, being more conservative, manage to remain in power but only at the expense of reduced encroachment. Thus, a prediction of the model is that power transitions are more likely when public beliefs are more optimistic.

The model also suggests that competition cannot fully undermine the reluctance of competing groups to reveal their private information. The reason is that the ruler is willing to make offers to a high-cost group if this allows him to extract information from his opponent. As a result, for low-cost groups the future value of being believed high-cost is high despite competition. In turn, this implies that to be willing to reveal information in the present, a low-cost group requires a sufficiently high transfer that even a high-cost group is ready to accept, undermining the possibility of learning in the first place. This is, of course, close to the “too much pooling result” from Laffont and Tirole (1988). In this context, it is the ability to exploit competition to undermine the ratchet effect in the future what limits the ability of the incumbent to do in the present.

The model also offers some insights about the question of which groups should be included
more often in ruling coalitions, and about the nature of coalitions themselves. For example, the simple intuition that the “cheapest” groups should always be included over the “expensive” ones proves to be wrong under some circumstances. In particular, when a ruler is unwilling to incur any risk of getting ousted, including expensive groups first is an effective way to exploit competition and force competing groups to reveal information. Supermajority coalitions are thus possible in equilibrium, but the rationale for them is not as in Groseclose and Snyder (1996), where non-minimum winning coalitions arise to make unilateral defections unprofitable. The model suggests that we should see supermajority coalitions as information extraction devices.

On the other hand, when a ruler is willing to incur some risk of being ousted, making initial offers to the group over which he is more pessimistic, turns out to be optimal too. The reason is that in case of a rejection by the first group being offered, facing a group which is ex-ante more likely to accept is cheaper when the ruler has to offer pivotality rents; this is, rents to compensate a group for not triggering a political transition. In fact, most of the strategic interaction between the incumbent and the potential coalition members is driven by the tension between what I call pivotality rents and informational rents. These will determine the relative cost of extracting information for the ruler, and the attractiveness of revealing information to the potential coalition members. Nevertheless, although it is attractive to make first offers to the relatively “more expensive” groups, in equilibrium this cannot be achieved for a large range of beliefs. The reason is an open set problem: if the ruler makes high offers to the group that looks expensive, this gives current incentives to the group looking less expensive to make support decisions such that in the future public beliefs about itself are as close as possible, but slightly more pessimistic, than beliefs about the competing politician.

The rest of the paper is organized as follows. Section 2 describes the benchmark model’s environment and introduces notation. For comparative purposes and to make intuitions about the different components of the model clear, the characterization of equilibrium behavior begins in section 3, looking at a model under incomplete information but in the absence of political competition. It shows the resilience of the ratchet effect, in that the incumbent can learn only if he is willing to incur the risk of losing power. Section 4 then looks at the model under political competition but with complete information. It shows that political competition is a powerful tool allowing a ruler to fully entrench and encroach by playing the different potential coalition members against each other. Section 5 subsequently develops the full model and presents the main results of the paper, and section 6 makes some concluding observations. The main proofs and mathematical derivations are left to the Appendices.
2 The Model

2.1 The Environment

Consider a discrete time, infinite-horizon, three player game. The set of players is \(\{r, a, b\}\), representing the ruler and two potential coalition members which I will henceforth call politician \(a\) and politician \(b\), respectively. Each politician is endowed with political power, meaning the ruler needs the political support of at least one politician to remain in power onto the next period. The ruler uses state resources, of which there is \(\mu\) available, to pay for political support, and is the residual claimant after paying for support. Every period the ruler publicly announces a politician he will make a first offer \(m^1_t\), such that \(m^1_t \leq \mu\). This politician must decide whether to support the ruler and take the transfer, or not to support him and expect a political transition. I denote \(D^i_t \in S = \{A, R\}\) to be the observable Acceptance or Rejection decision of politician \(i\), and let \(\sigma^i_t \in [0, 1]\) be his chosen probability of acceptance. After observing the decision of the first politician being offered, the ruler decides which offer \(m^2_t\) to make (if any) to the remaining politician, who must also decide whether to accept it or reject it. If neither politician accepts, the ruler is ousted from office, and a payoff \(\phi\) is divided among them\(^1\). The assumption that offers are made sequentially has the purpose of abstracting from any coordination issues, which would, of course, give rise to multiple equilibria, and are not at the heart of the main mechanism the model intends to highlight. Moreover, assuming the uninformed ruler is the one making the offers to the informed politician limits the signaling incentives of politicians by restricting information transmission to depend upon their acceptance or rejection decision exclusively. This limits the impact that specific assumptions about beliefs following probability-zero events have on the set of equilibria (see Fudenberg and Tirole (1991, p. 404)).

There is heterogeneity in the cost of supplying political support; each politician \(i \in \{a, b\}\) has a privately known type \(\theta^i \in \Theta = \{\theta, \overline{\theta}\}\), which measures how dependent upon State resources he is to mobilize political support. A politician type \(\theta^j\) requires at least \(\theta^j\) worth of transfers from public resources to be able to mobilize his political base to support the ruler. This is a dynamic incomplete-information game, where a state can be identified by the current public beliefs about types, \(\Omega_t = (\omega^a_t, \omega^b_t) \in [0, 1]^2\). I define \(\omega^i_t\) as the posterior probability that politician \(i\) is type \(\theta^i\). This is, the information of player \(i \neq r\) at time \(t\) is \(\Omega^i_t \equiv \Omega_t \cup \{\theta^i\}\), while the ruler’s information is just \(\Omega_t\). The state-space is thus \([0, 1]^2\).

An (ex ante) public history \(h^t = (\{D^i_k, m^i_k\}_{i=a,b}, \tau_k, \Omega_k)_{k=0}^{t-1}, \Omega_t\) can be defined as the vector of action profiles of all players up to time \(t\), plus the sequence of states up to \(t\). \(\tau_k \in I\) denotes the identity of the politician to whom the first offer is made in period \(t\). Let \(H^t\) be the set

\(^1\phi\) can be thought of as the present value of state resources left after the (possibly costly) power transition has taken place.
of possible time-$t$ histories, and $H = \bigcup_{i=0}^{\infty} H^i$ be the set of all possible histories. I define the initial history as the prior at $t = 0$, $H^0 \equiv \Omega_0$. A (possibly) mixed strategy for the ruler will be a triple of mappings $\Upsilon : H \to [0, 1]$, $m^t : H \to [0, \mu]$, and $m^{\sim t} : H \times S \to [0, \mu]$. where $\Upsilon(h)$ denotes the probability that the ruler makes the first offer to $a$ after history $h$, $m^t(h)$ denotes the offer to the politician being made a first offer, and $m^{\sim t}(h, D)$ denotes the offer to the politician being made a second offer, which might depend on the support decision of the first offered politician. A (possibly) mixed strategy for politician $i$ will be a mapping $\sigma^i : H \times [0, \mu] \times I \times S \times \Theta \to [0, 1]$, where $\sigma^i(h, m^i, t, D^{\sim t}; \theta) = Pr(D^i = A|h, m^i, t, D^{\sim t}; \theta)$ denotes the probability that politician $i$ type $\theta$ accepts an offer $m^i$ after public history $h$, when the first politician being made an offer in the current period is $t$, and $D^i$ is $i$’s support decision (of course, if $i = \nu$, the strategy does not depend on $D^{\sim i}$).

2.1.1 Timing and Definition of Equilibrium:

The specific timing of events is as follows:

1. The ruler chooses a politician (possibly with a mixed strategy $\Upsilon$) to make a first offer (label this politician $p^1 \in \{a, b\}$), and the amount of such offer $m^1$.

2. If $m^1 \geq \theta^1$, $p^1$ decides whether to accept or reject (possibly with a mixed strategy) $\sigma^1$, his decision $D^1 \in \{A, R\}$ is publicly observed, and the transfer is made.

3. Beliefs about $p^1$ are updated according to Bayes’ rule. The ruler decides the amount $m^2$ to offer to the remaining politician $p^2$.

4. If $m^2 \geq \theta^2$, $p^2$ decides whether to accept or reject (possibly with a mixed strategy) $\sigma^2$, his decision $D^2 \in \{A, R\}$ is publicly observed, and the transfer is made.

5. Beliefs about $p^2$ are updated according to Bayes’ rule.

6. If at least one politician accepts ($((D^1, D^2) \in \{(A, R), (R, A), (A, A)\})$, the ruler remains in power onto the next period. If no politician accepts ($((D^1, D^2) = (R, R)$), the ruler is ousted, and a rent $\phi$ is divided evenly among them.

The analysis will restrict attention to Markov Perfect Bayesian Equilibria (MPBE) of the game, where at any stage, strategies are dependent only upon the current state of the game and possibly any actions taken during the stage game (i.e., only on payoff-relevant variables), and beliefs are updated according to Bayes’ rule whenever possible. In this model, all previous distributions of State rents ($\{m^i_{k} \}_{k=0}^{t-1}$), all previous decisions of politicians ($\{\sigma^i_{k} \}_{k=0}^{t-1}$), all previous beliefs about politicians’ types ($\{\Omega^i_{k} \}_{k=0}^{t-1}$), and all previous identities of the first politician being offered ($\{i^k_{k=0} \}_{k=0}^{t-1}$) are irrelevant to current payoffs. The only payoff-relevant variables are current beliefs, current offers, and current support decisions. Thus, a Markov strategy for the ruler will be a triple of mappings $\Upsilon : [0, 1]^2 \to [0, 1]$, $m^t : [0, 1]^2 \to \mathbb{R}^2_+$.
and \( m^*: [0,1]^2 \times \{A, R\} \rightarrow \mathbb{R}^2_+ \). A Markov strategy for politician \( i \) will be a mapping \( \sigma^i: [0,1]^2 \times [0,\mu] \times I \times S \times \Theta \rightarrow [0,1] \). More formally,

**Definition.** An equilibrium of this game consists of functions \( \Upsilon(\Omega), m^*(\Omega), m^*(\Omega, D^i), \sigma^a(\Omega, m^a, \iota, D^b; \theta), \sigma^b(\Omega, m^b, \iota, D^a; \theta) \), specifying the ruler’s and politicians’ actions under any beliefs \( \Omega \) (the current state), such that neither has a profitable deviation at any \( \Omega \), and beliefs are updated according to Bayes’ rule whenever possible, evolving according to equation (1). Off-the-equilibrium-path actions are assumed to be uninformative, this is, when Bayes’ rule cannot be applied, public beliefs are assumed to remain unchanged.

### 2.1.2 Learning, Beliefs and the state-space:

Beliefs are rationally updated according to Bayes’ rule whenever possible. Equilibrium beliefs evolve as a function of the observed decisions and equilibrium strategies of politicians, and of the offer vector chosen by the ruler. To simplify notation, I will denote the low-cost and high-cost type politicians’ strategies by \( \sigma \) and \( \bar{\sigma} \) respectively, and will omit the identity of the first politician being offered from the arguments, unless it is unclear from the context.

By definition, politician \( i \)’s strategy is the conditional probability of acceptance given his type: \( Pr(A_i|\Omega, m, D^{-i}, \theta) = \sigma^i(\Omega, m, D^{-i}) \). Hence, from Bayes’ rule the transition kernel for equilibrium beliefs can be expressed as a function of the politicians’ strategies:

\[
\omega_{t+1}^i(D^i; \Omega, m) = \begin{cases} 
\frac{\sigma^i(\Omega, m, D^{-i}) \omega^i_t}{\sigma^i(\Omega, m, D^{-i}) \omega^i_t + \bar{\sigma}^i(\Omega, m, D^{-i}) \omega^i_t} & \text{if } D^i = A \\
\frac{(1-\sigma^i(\Omega, m, D^{-i}) \omega^i_t + \bar{\sigma}^i(\Omega, m, D^{-i}) \omega^i_t) \omega^i_t}{1-\sigma^i(\Omega, m, D^{-i}) \omega^i_t + \bar{\sigma}^i(\Omega, m, D^{-i}) \omega^i_t} & \text{if } D^i = R
\end{cases}
\]  

(1)

As a remark, notice that offers over which both types of politician pool will leave beliefs unchanged. As a result, if the ruler wants to make an informative offer (an offer that induces some extent of separation), he has to make offers that are in principle risky, since they must allow for a positive probability of being rejected. As noted above, the state-space of this game is the set \([0,1]^2\), depicted below. As a convention, throughout the paper beliefs about \( a \) will be plotted in the x-axis, and beliefs about \( b \) will be plotted in the y-axis. In Figure 1 below I have plotted a 45 degree line on the state-space. This is because the dynamics of the game will depend closely on relative beliefs, this is, on which politician is deemed more likely to be low cost at any point in time. Under beliefs \( \Omega \) below the 45 degree line, \( \omega^a \leq \omega^b \), so politician \( a \) is deemed more likely to be low cost. Moreover, the Markov restriction implies that equilibrium strategies must be symmetric around the 45 degree line.

---

1 believe this is the most parsimonious assumption to make about off-path beliefs, since, as it will become clear below, for different strategy profiles and beliefs, upwards or downwards incentive constraints might bind. As a result, other assumptions about off-path beliefs might imply "punishments" for a given type under certain strategies and states but not under others.
I proceed below with a general recursive characterization of the problem, which will be used throughout the paper.

2.2 Analysis

2.2.1 The Ruler’s problem:

Define $W^+(\Omega)$ to be the ex-ante equilibrium value for an incumbent ruler under beliefs $\Omega$. Define $V^+(\Omega, t)$ to be the interim equilibrium value for the ruler when the state is $\Omega$ and he the state is $\Omega$ and he decided to make a first offer to politician $t$, and $v^+(\bar{\Omega}, D^1)$ to be the interim equilibrium value for the incumbent ruler after observing $p^1$’s decision $D^1$. At this stage the ruler must decide which offer to make to $p^2$. The problem is

$$v^+(\bar{\Omega}, D^1) = \mu - m^1 I\{D^1 = A\} + \max_{m^2} \{-m^2 Pr(D^2 = A|D^1, m^2, \bar{\Omega}) \}
$$

+ $I\{D^1 = A\} \sum_{j \in \{A, R\}} Pr(D^2 = j|A, m^2, \bar{\Omega}) W^+(\bar{\Omega}(j, m^2; \bar{\Omega}))+I\{D^1 = R\} Pr((D^2 = A|R, m^2, \bar{\Omega}) W^+(\bar{\Omega}(A, m^2; \bar{\Omega}))

(2)

where of course $Pr(D^2 = A|D^1, \bar{\Omega}, m^2) = \omega^2 \bar{\sigma}^2(\bar{\Omega}, m^2, D^1) + (1 - \omega^2)\bar{\sigma}^2(\bar{\Omega}, m^2, D^1)$, and $Pr(D^2 = R|D^1, \bar{\Omega}, m^2) = \omega^2(1 - \bar{\sigma}^2(\bar{\Omega}, m^2, D^1)) + (1 - \omega^2)(1 - \bar{\sigma}^2(\bar{\Omega}, m^2, D^1))$. Equation (2) takes into account that the ruler looses power if $p^2$ rejects after $p^1$ rejected. Of course, $\bar{\Omega} \equiv \Omega'(D^1, m^1; \Omega)$ denotes interim posterior beliefs after $p^1$ has made his decision. Here I make explicit that the evolution of beliefs depends not only on the support decision of
politicians, but also on the offer made and on prior beliefs. The interim value for the ruler when making the first offer can be expressed recursively as

$$V'(\Omega, i) = \max_{m^1} \{-m^1 Pr(D^1 = A|m^1, \Omega)$$

$$+ \sum_{j \in \{A,R\}} Pr(D^1 = j|m^1, \Omega) v'(\Omega(j, m^1; \Omega), j)\}$$

(3)

with $Pr(D^1 = A|m^1, \Omega) = \omega^1 \theta^1(m^1, \Omega) + (1 - \omega^1)\bar{\theta}^1(m^1, \Omega)$. Finally, the ex-ante value for the ruler is given by

$$W^r(\Omega) = \max_{\Upsilon \in [0,1]} \Upsilon V^r(\Omega, a) + (1 - \Upsilon) V^r(\Omega, b)$$

(4)

Of course, this trivially implies that the ruler will chose $\Upsilon = 1$ if $V^r(\Omega, a) > V^r(\Omega, b)$, $\Upsilon = 0$ if $V^r(\Omega, a) < V^r(\Omega, b)$. I will assume the ruler sets $\Upsilon = 1/2$ if indifferent.

### 2.2.2 The Politician’s Problem:

Define $V^i_\theta(D^1, m^i, \Omega)$ to be the interim value for politician $i$ type $\theta$ when the state is $\Omega = (\omega^a, \omega^b)$ and he is offered $m^i$. Of course, if $i$ is offered first, his strategy does not depend on $D^1$. The politician must decide a probability of accepting a given offer, by trading-off the current payoff he might get, the informational consequences his decision might have on future beliefs, the possibility of ousting the ruler, and the expected response from the competing politician. If politician $i$ accepts the offer today, he gets the current payoff $m^i - \theta^i$, and there is no political transition. Otherwise he gets no rents in the present. The ruler might be retained or not depending on the support decision of his competing politician $\sim i$. In the latter case, if $\sim i$ also rejects the offer, the ruler is ousted and each politician gets half of an exogenous payoff from having the ruler ousted. It is also important to notice that $i$’s strategy is a function not only of his beliefs about politician $\sim i$, but also of the public beliefs about his own type. This is due to the competition among political entrepreneurs. A low cost politician $i$ benefits from diffuse public beliefs about his type, but to the extent that he might look “expensive” relative to $\sim i$, such diffuse beliefs about his own type might hurt. Thus, optimal behavior should also take into account public information about one’s own type. Finally, the politician also must take into account that his current decision, through the impact it will have on posterior beliefs, will affect the likelihood that he is offered first or second in the future. The analysis will rely heavily on studying the incentive constraints faced by politicians at any point in time, so here I state the most general for that such an
incentive constraint will have. A type-θ politician a will be willing to accept a first offer $m^1$ only if

$$m^1 - \theta + Pr(D^b = A|A, \tilde{\Omega}, m^2)\delta V^a_\theta(\omega^a(A), \omega^b(A)) + Pr(D^b = R|A, \tilde{\Omega}, m^2)\delta V^a_\theta(\omega^a(A), \omega^b(R)) \geq Pr(D^b = A|R, \tilde{\Omega}, m^2)\delta V^a_\theta(\omega^a(A), \omega^b(A)) + Pr(D^b = R|R, \tilde{\Omega}, m^2)\delta \frac{1}{2}\phi$$

(5)

where $m^2$ depends on interim beliefs $\tilde{\Omega}$ and $D^a$. Equation (5) above highlights that in a MPBE the support decision of a politician will depend on the ruler’s subsequent offer, on the supporting strategy of the competing politician, and on the evolution of public beliefs following acceptance or rejection decisions. An analogous expression can be obtained for the second politician to be offered, which I omit here to save on space.

3 No Competition under Incomplete Information

I begin by studying a simplified environment in the absence of political competition, which helps to highlight the role that incomplete information alone plays in the entrenchment and encroachment dynamics. Thus, assume for now there is a ruler $r$ and only one other politician $p$ with private information about his cost type $\theta \in \{\theta, \bar{\theta}\}$, $\theta < \bar{\theta}$. There is a prior $\omega$ about $p$’s type, where $\omega$ is the probability that $p$ is low-cost. The ruler needs the support of this politician to remain in power unto the following period. If $p$ does not support, the ruler is ousted from power, and $p$ takes his place as ruler. I assume in such case $p$ gets an exogenous payment $\phi$ and the game ends.

Let $\sigma(\omega, m)$ be the probability that a low-cost type politician supports the ruler when public beliefs are $\omega$ and the ruler has offered him $m$, and $\sigma(\omega, m)$ be the corresponding probability for a high-cost type. Given that $\omega = 0$ and $\omega = 1$ are absorbing states, I first start by characterizing the equilibrium values for the ruler and the politician under each of them. Throughout I will assume the discount factor is large.$^3$

**Proposition 1.** Let $\phi < \frac{\delta \bar{\theta} - \theta}{\delta (1 - \delta)}$, and assume $\omega = 1$ (it is common knowledge that the politician is low-cost). The unique MPE takes following form: Every period, the ruler offers $m = \theta + \delta (1 - \delta)\phi$ to the politician, and the politician supports with probability 1.

$^3$The results for the full model developed below will require that $\delta \geq \theta / \bar{\theta}$, so that the larger the gains from information extraction, the larger the range of discount factors that will be able to support the equilibria characterized in the paper.
Proof. See Appendix 1. □

The equilibrium for $\omega = 0$ is of course identical to the above, replacing $\theta$ for $\bar{\theta}$ everywhere. The parameter restriction in Proposition 1 is the necessary and sufficient condition for the ruler to prefer remaining in power than taking all state resources and being ousted when $\omega = 0$. It sets an upper bound on $\phi$ which depends non-trivially on the discount factor. A higher $\delta$ makes the ruler more willing to remain in power in the future, but also makes the politician’s value of ousting the ruler higher. I will maintain this parameter restriction throughout. Notice that if we think of state resources $\mu$ as large relative to the cost of providing political support, this restriction on discount factors is a weak one. The equilibrium values for the ruler are $W'^r(0) = \frac{\nu - \bar{\theta}}{1 - \delta} - \delta \phi$ and $W'^r(1) = \frac{\nu - \theta}{1 - \delta} - \delta \phi$, and the equilibrium values for the politician are $V'(1) = V'(0) = \delta \phi$.

Under complete information and no political competition, the dynamics are straightforward. The ruler is willing to remain in power, and he offers a transfer to the politician just high enough to make him willing to support every period. This transfer consists of the politician’s type, plus a pivotality rent that compensates him for not triggering a political transition. The ruler remains in power with probability 1, so he is fully entrenched, but his encroachment is limited by the pivotal power of the supporter. To characterize the equilibrium for interior $\omega$, I make use of the following key result which will be used throughout the paper.

**Lemma 2.** The politician’s equilibrium strategies must be decreasing in type: $\sigma(m, \omega) \leq \sigma(m, \omega)$. Moreover, $\sigma(m, \omega) \in (0, 1)$ implies $\sigma(m, \omega) = 0$.

Proof. See Appendix 1. □

The intuition for Lemma 2 is straightforward. In equilibrium, any given offer that a high-cost politician can accept, must also be feasible for a low-cost politician, and must leave him a strictly larger surplus. Thus, if the incentive constraint for acceptance is satisfied for the high-cost type, it must be slack for the low-cost type. The Lemma provides an even stronger result; if a low-cost politician is mixing, the high-cost politician must be rejecting. The reason is that for the low-cost politician to mix, he must be exactly indifferent between acceptance and rejection, which means his incentive constraint will be binding. If the low-cost type’s incentive constraint is binding for a given offer $m$, the high-cost type’s incentive constraint for must not be satisfied. Lemma 2 restricts the possible types of continuation equilibria (after the ruler has made an offer $m$), which must take one of the following forms:

1. Pooling on rejection (PR): $\sigma(m, \omega) = \bar{\sigma}(m, \omega) = 0$.  

107
2. Low-cost type semi-separating (LSS): $\sigma(m, \omega) \in (0, 1)$ and $\overline{\sigma}(m, \omega) = 0$. Such an equilibrium implies posterior beliefs must be $\omega(A) = 1$, $\omega(R) < \omega$.

3. Separating (S): $\sigma(m, \omega) = 1$, $\overline{\sigma}(m, \omega) = 0$. Such an equilibrium implies posterior beliefs must be $\omega(A) = 1$, $\omega(R) = 0$.

4. Pooling on acceptance (PA): $\sigma(m, \omega) = \overline{\sigma}(m, \omega) = 1$. Such an equilibrium implies posterior beliefs must be $\omega(A) = \omega(R) = \omega$.

5. High-cost semi-separating (HSS): $\sigma(m, \omega) = 1$, $\overline{\sigma}(m, \omega) \in (0, 1)$. Such an equilibrium implies posterior beliefs must be $\omega(A) > \omega$, $\omega(R) = 0$.

The nature of belief updating implied by rationality together with the fact that politician’s strategies are decreasing in type, and moreover, that a mixing strategy for a low-cost type implies a high-cost type must be rejecting, imposes strong restrictions on the possible evolution of beliefs in a Bayesian equilibrium and on the possible continuation equilibria of the game. Figure 2 graphically depicts the possible continuation equilibria (these results are proven in Lemmas A1-1, A1-2, A1-3 and A1-4 in Appendix 1). Both types pool on acceptance when the offer is $m \geq \bar{\theta} + \delta(1 - \delta)\phi$, high-cost types reject any offer below $\bar{\theta}$, and low-cost types reject any offer below $\bar{\theta}$. For offers $m \in [\theta, \bar{\theta} + \delta(1 - \delta)\phi)$, both types of politician must pool on rejecting. Finally, for offers $m \in [\bar{\theta} + \delta(1 - \delta)\phi, \bar{\theta} + \delta(1 - \delta)\phi)$, types fully separate, so that a low-cost politician accepts with probability 1, while a high-cost politician rejects.

Figure 2: Continuation equilibria under no competition

In an environment without political competition, a ruler who requires the acquiescence of a politician to remain in power, and who has the possibility of extracting information about his cost of providing support, has three options; he might make a high enough offer which both
low and high cost politicians accept (but which will not allow him to learn about his type),
he might make a low offer which both politicians reject and which triggers a power transition,
or he might make an intermediate offer which separates the behavior of low and high-cost
politicians (but which implies there will be a positive probability that he is ousted from
power). The value of the latter strategy depends on what the beliefs about the politician’s
type are, and will play a central role for the equilibrium dynamics. More pessimistic beliefs
imply a higher likelihood that a given separating offer will be rejected. For \( \phi < \frac{\delta \mu - \bar{\theta}}{\delta (1 - \delta)} \) we
already know that the ruler prefers making a large offer and forgo learning over being ousted
from power for sure. It remains to study the range of beliefs for which he might be willing
to incur risk of a power transition, hoping to learn he faces a low-cost politician.

First, observe that for any offer \( m \in [\theta + \delta (1 - \delta) \phi, \bar{\theta} + \delta (1 - \delta) \phi] \) the low-cost politician
supports with probability 1 while the high-cost politician rejects with probability 1. Thus,
if the ruler wants to induce separation he will always offer \( m = \theta + \delta (1 - \delta) \phi \). The value of
pursuing this strategy under belief \( \omega \) is

\[
W^r(\omega) = \mu - \omega [\theta + \delta (1 - \delta) \phi] + \delta \omega W^r(1) = \mu + \omega \left[ \frac{\delta \mu - \theta}{1 - \delta} - \theta - \delta \phi \right]
\]

which is a linear and increasing function of \( \omega \). The value of pursuing the safe no-learning
strategy is just \( W^r(0) \), which implies the ruler prefers to incur positive risk of being ousted
iff \( \omega + \omega \left[ \frac{\delta \mu - \theta}{1 - \delta} - \theta - \delta \phi \right] > \frac{\mu - \bar{\theta}}{1 - \delta} - \delta \phi \). Solving for \( \omega \),

\[
\omega > \frac{\delta \mu - \bar{\theta} - \delta (1 - \delta) \phi}{\delta \mu - \theta - \delta (1 - \delta) \phi} \equiv \tilde{\omega}(\delta) \in (0, 1)
\]

The cutoff value of beliefs at which the ruler starts incurring risk is an increasing function of
\( \delta \) as long as \( \delta > \frac{1}{2} \left( \frac{\phi - \mu}{\phi} \right) \). If \( \phi < \mu \), the cuttof is increasing in \( \delta \) for any \( \delta \). As the ruler becomes
more patient, the region of beliefs where he cannot learn widens, because it becomes more
costly to incur the risk of not being the ruler in the future. Nonetheless, if \( \phi > \mu \), there is a
region of small values of \( \delta \) for which higher discount factors can reduce the cuttof. The reason
is that when \( \phi \) is large it is too attractive for a politician to trigger a power transition. As \( \delta \)
increases starting from a low level, the ousting option value for the politician also increases,
making the value of learning for the ruler increase, and thus making him more willing to
incur risk. Also, observe that the value for either type of politician is \( V_\theta(\omega) = \delta \phi \). Having
characterized the continuation equilibria for given offer \( m \), the following proposition follows:

**Proposition 3.** Let \( \omega \in (0, 1) \). The unique MPBE takes following form: For \( \omega < \tilde{\omega}(\delta) \) the
ruler makes an offer \( m = \bar{\theta} + \delta (1 - \delta) \phi \) which both types of politician accept with probability

\[\text{Notice that the assumption } \delta \mu - \bar{\theta} > \delta (1 - \delta) \phi \text{ implies the term in square brackets is positive}\]
1. For $\omega \geq \tilde{\omega}(\delta)$ the ruler makes a separating offer $m = \theta + \delta(1 - \delta)\phi$, which a low-cost politician accepts with probability 1 and a high-cost politician rejects with probability 1.

Proof. See Lemma 2 above and Appendix 1.

Figure 3 below presents the equilibrium value function for the ruler. It illustrates the ratchet-effect dynamics of the no-competition game. For a large range of beliefs, the ruler is unable to extract any information, and is only capable of doing so at the expense of incurring the risk of losing power with positive probability. Observe that in the pooling region ($\omega < \tilde{\omega}$) the ruler is offering $m = \bar{\theta} + \delta(1 - \delta)\phi$ and there is no learning, so that a both types are receiving a pivotality rent equal to $\delta(1 - \delta)\phi$, to compensate them for not triggering a political transition, while the low-cost type is also receiving an informational rent worth $\bar{\theta} - \theta$. In the separating region, the low-cost politician loses the ability to extract an informational rent, but only because the ruler is willing to put at stake his permanence in power. A pessimistic enough ruler will not be able to undermine the ratchet effect, while in the learning region he is ousted with probability $1 - \omega$. Interestingly, a power transition only happens for high values of $\omega$, this is, when the ruler is relatively optimistic about the politician’s type.

Figure 3: Ruler’s Equilibrium Value under no competition

4 Competition under Complete Information

Now I look at an environment with political competition in the coalition formation game, but where politicians’ types are common knowledge. In this context, trivially the ruler fully takes advantage of the competitive nature of the game to take politicians down to their participation constraints. Hence, encroachment and entrenchment are complete. Because
the complete information game is subsumed in the incomplete information game for \( \Omega \in \{(0, 0), (0, 1), (1, 0), (1, 1)\} \), and revelation of types will take beliefs to one of these states (which are absorbing), I start by characterizing the MPE under them.

**Proposition 4.** The complete information game when \( \Omega = (1, 1) \) has a unique MPE of the following form: Every period the ruler picks one politician at random (\( p^1 \)) to make him a first offer \( m^1 = \theta \), which \( p^1 \) accepts. Then the ruler makes no second offer to the remaining politician \( p^2 \). Off-the-equilibrium path if \( p^1 \) rejects, the ruler makes a second offer \( m^2 = \theta + \frac{1}{2} \delta \phi \) to \( p^2 \) which \( p^2 \) accepts (the complete formal description of equilibrium strategies can be found in Appendix 2).

Of course, the unique equilibrium for \( \Omega = (0, 0) \) is identical just replacing \( \theta \) for \( \tilde{\theta} \) above.

**Proof.** See Appendix 2.

The ruler is completely entrenched (remains in power with probability 1) and fully encroached (leaves no rents in any period on the equilibrium path). Proposition 4 highlights that the ruler’s ability to make an offer which leaves no positive rents to any of his supporters relies on the fact that he can credibly threaten the first politician he makes an offer to, that after a rejection decision, an offer will be made to the remaining politician, which he will accept for sure. Thus, although the equilibrium coalition is minimum winning, \( p^1 \) is not pivotal in equilibrium; the threat of competition makes his pivotality worthless. The equilibrium value for the ruler is \( W^r((1, 1)) = \frac{\nu - \theta}{1 - \delta} \), and for both politicians it is \( V^a((1, 1)) = V^b((1, 1)) = 0 \). Under \( \Omega = (0, 0) \) the value for the ruler is \( W^r((0, 0)) = \frac{\nu - \theta}{1 - \delta} \), and for both politicians it is \( V^a((0, 0)) = V^b((0, 0)) = 0 \). Of course, the ruler does strictly better when facing two low-cost, instead of two high-cost politicians. Now I look at the equilibrium under asymmetric types:

**Proposition 5.** The complete information game when \( \Omega = (0, 1) \) has a unique MPE of the following form: Every period, the ruler makes a first offer \( m^1 = \theta \) to \( b \) (the politician whose cost is known to be low), which \( b \) accepts. Then the ruler makes no second offer to \( a \) (the politician whose cost is known to be high). Off-the-equilibrium-path if \( b \) rejects, the ruler makes a second offer \( m^2 = \tilde{\theta} + \frac{1}{2} \delta \phi \) to \( a \) which \( a \) accepts.

Of course, the unique equilibrium for \( \Omega = (1, 0) \) is identical just changing \( a \) for \( b \) above.

**Proof.** See Appendix 2.
The ruler is completely entrenched and fully encroached on the equilibrium path. Again, the ability to make a minimum winning coalition which drives the included politician down to his cost relies on the ruler’s ability to credibly threaten making an acceptable offer to the competing politician in case the politician who is offered first does not accept. The credibility of this threat is based on the fact that types are common knowledge. For the ruler, political competition is a much better environment compared to the situation without political competition and complete information. Recall that when facing a high-cost politician, the ruler had to offer \( m = \bar{\theta} + \delta(1 - \delta)\phi > \bar{\theta} \). Compared to the no-competition case, here competition makes both politicians effectively non-pivotal, and the ruler never has to give pivotality rents. The second politician being offered is not pivotal because the first one has already accepted, and the first politician is not pivotal either because he knows that after rejecting, the second politician to be offered will accept for sure. Key for the ability of the ruler to make the first politician offered non-pivotal is, thus, the belief that he can effectively make a second offer that will be accepted for sure after a rejection.

The equilibrium value for the ruler is \( W^r((0,1)) = \frac{\nu - \theta}{1-\delta} \), and for both politicians it is \( V^a((0,1)) = V^b((0,1)) = 0 \). Of course, the equilibrium is symmetric for \( \Omega = (1,0) \). The results of Propositions 4 and 5 show that the ruler is actually indifferent between states \( \Omega = (0,1) \), \( \Omega = (1,0) \), and \( \Omega = (1,1) \), whereas he strictly prefers any of those states to state \( \Omega = (0,0) \).

5  Competition under Incomplete Information

5.1 One low-cost Politician

Now I proceed to study the coalition-formation game under both competition and incomplete information. In the complete-information game analysis of section 4, the equilibrium value for the ruler at \( \Omega = (1,0) \) and at \( \Omega = (1,1) \) was the same, namely \( W^r((1,0)) = W^r((1,1)) = \frac{\nu - \theta}{1-\delta} \). The ruler does as well when he knows both politicians are low-cost than when he knows one of them is low-cost and the other is high-cost. Given that the ruler only needs the support of one other politician to remain in power, does this imply that for any beliefs \( \Omega \in \{(1, \omega) : \omega \in (0,1)\} \), where the ruler knows one of the politicians is low-cost, his value is independent of beliefs about the competing politician? The analysis of section 3 already suggests this will not be the case, because encroachment incentives create a commitment problem for the ruler. Given beliefs \( \omega \) about \( b \) (the politician whose type is not common knowledge), a ruler would benefit from being capable of committing to learn nothing about \( b \) with a pooling offer after \( a \) has rejected, just as it happens when types are common knowledge. In that setting, \( a \) knows that off path after a rejection, the ruler will credibly make an offer.
to $b$ high enough that $b$ will accept with probability 1. But when beliefs about $b$'s type are interior, after $a$ has rejected the ruler might still want to make a low enough, separating offer to $b$ that is informationally valuable. The incentives to pursue this strategy will depend on how large informational rents are relative to pivotality rents. This possibility would actually give $a$ incentives to reject, in the hope that $b$ is a high-cost type who will reject the offer made by the ruler (or even a low-cost type who is not accepting with probability 1).

The ruler’s temptation for information extraction following a rejection gives $a$ an option value of rejecting, making him, in equilibrium, more expensive. The no-competition game already illustrated that for high values of $\omega$ the ruler will indeed be willing to incur in equilibrium risk of being ousted from power. Of course, such a lack of commitment capacity by the ruler, by making $a$’s outside value higher, implies that the ruler’s value need not be the same at $\Omega = (1, \omega)$ than at $\Omega = (1, 0)$ or $\Omega = (1, 1)$. Indeed, the ruler’s equilibrium value will fall discretely at the value of beliefs where the commitment problem arises.

The inability to commit will only be binding for large enough values of $\omega$, where it is likely enough that $b$ is a low-cost politician. As a result, the ruler will find it attractive to induce a non-pooling equilibrium through a relatively low offer, just as in the no-competition case. This cutoff will be given by the value of beliefs for which, at the interim stage following a rejection, the ruler is indifferent between making a pooling offer and the one-shot deviation consisting of a separating offer to the remaining politician:

$$v^*(\Omega)|_{\text{Separate}} = \mu + \bar{\omega}[\delta(\mu - \theta) \theta = \delta(1 - \delta) - \delta(1 - \delta)\frac{1}{2} \phi] = \mu - \bar{\theta} - \delta(1 - \delta)\frac{1}{2} \phi + \delta(\mu - \theta) \theta = v^*(\Omega)|_{\text{Pool}} \Leftrightarrow$$

$$\bar{\omega} = \frac{\delta(\mu - \theta) \theta - \delta(1 - \delta)\frac{1}{2} \phi, \delta(\mu - \theta) \theta - \delta(1 - \delta)\frac{1}{2} \phi} \in (0, 1) \quad (8)$$

At the interim stage following a rejection, a separating offer is accepted with probability $\bar{\omega}$, in which case the ruler learns he is facing a low-cost type and gets fully entrenched and encroached. It is rejected with the complementary probability, in which case the ruler is ousted. On the other hand, by making a pooling offer the ruler forgoes any learning. This lack of commitment is purely driven by encroachment incentives, because a non-pooling offer is attractive only because it implies a lower transfer has to be made in the event that $b$ accepts. Throughout I will focus on the equilibrium dynamics for large $\delta$. This is the case for which the ratchet effect is more likely to limit entrenchment, given that for low discount factors politicians are trivially willing to reveal their type for any current positive transfer.

Notice that $\bar{\omega}$ is lower the larger is $\bar{\theta}$ relative to $\theta$, so that when the value of learning is very high, the range of beliefs for which the ruler faces a commitment problem widens. Proposition
Proposition 6. Assume \( \Omega \in \{(\omega^a, \omega^b) : \omega^a = 1, \omega^b \in (0, 1)\} \). The MPBE can be characterized as follows. For \( \omega \leq \bar{\omega} \) the ruler makes a first offer \( m^1 = 0 \) to \( a \) (the politician whose type is known to be low) which \( a \) accepts, and then the ruler makes no second offer to \( b \) (the politician whose type is not common knowledge). Off-the-equilibrium-path if \( a \) rejects, the ruler makes a pooling second offer \( m^2 = \bar{\theta} + \delta \frac{1}{2} \phi \) to \( b \), which both types of \( b \) accept. For \( \omega > \bar{\omega} \) the ruler makes a first separating offer \( m^1 = \bar{\theta} + \frac{1}{2} \delta \phi \) to \( b \) which a low-cost \( b \) accepts with probability 1 and a high-cost \( b \) rejects with probability 1. After a support decision is observed the ruler makes no offer to \( a \), and if a rejection decision is observed the ruler makes a second offer \( m^2 = \theta + \frac{1}{2} \delta \phi \) to \( a \), which \( a \) accepts. The cutoff value \( \bar{\omega} \) is given by \( \bar{\omega} = \frac{\delta_{\mu-\theta}-(1-\delta)\bar{\theta}-\delta_{1-\delta}\frac{1}{2} \phi}{\delta_{\mu-\theta}-(1-\delta)\bar{\theta}-\delta_{1-\delta}\frac{1}{2} \phi} \).

Of course, the equilibrium is symmetric for \( \Omega \in \{(\omega^a, \omega^b) : \omega^a = 1, \omega^a \in (0, 1)\} \).

Proof. See Appendix 3.

Figure 3 below depicts the dynamics implied by proposition 6 for states \( \Omega = (1, \omega) \) or \( \Omega = (\omega, 1) \).

Figure 3: Equilibrium Dynamics at \( \Omega = (1, \omega) \) and \( \Omega = (\omega, 1) \) boundaries:

The ruler's ability of threatening with inducing \( b \) to pool off-the-equilibrium path undermines any incentives for \( a \) not to support the ruler in the first place. Nevertheless, recall from section 3 that when facing one politician (analogous to the off-path subgame after which \( a \)
has rejected), the ruler might have incentives to test a politician whose type is not known, when beliefs are sufficiently high. If off the equilibrium path the ruler cannot commit to induce a pooling equilibrium, his ability to drive \(a\) down to his cost will be undermined. In this sense, having sure information, even if it is negative (as knowing that \(b\) is high-cost) is better for the ruler than having uncertain information (as believing that \(b\) is low cost with high probability), because uncertain information creates a commitment problem.

For relatively low values of \(\omega\) there is no commitment problem. When deciding which politician to make a first offer when one of them is known to be low-cost, the choice is unambiguously clear; by making an offer to the known-type politician the ruler is able to buy his support with an offer of only his (low) cost. Moreover, he does not need to give any rents because the credible threat of a pooling offer to his competitor after a rejection, makes rejecting worthless. In contrast, making an offer first to the imperfectly known-type politician either requires a high offer on which both types pool, or a low offer which a high-cost politician rejects. This makes the known-cost-type politician pivotal on the equilibrium path, and hence requires a larger transfer to achieve his support. As long as the ruler can commit to offering a pooling transfer after a rejection, it is always strictly better to make a first offer to the low-cost known-type politician.

For high enough \(\omega\) the ruler loses his ability to commit not to learn. He can still make first an offer to the known low-cost type politician which includes an off-path pivotality rent, or he can first make a separating offer to the unknown type politician, and only after a rejection, make an offer to the known-cost type politician to gather his support for sure. It turns out these two strategies are payoff equivalent to the ruler. The assumption that the ruler prefers learning when indifferent implies he induces separation. Finally observe that for this set of beliefs there is never, on or off the equilibrium path, a supermajority coalition (a situation where both \(a\) and \(b\) support). This is because in the range for which the ruler has no commitment problem the ruler can achieve the maximum payoff without having to learn \(b\)'s type. For the range where he faces a commitment problem separation obtains, but after a type-revealing support decision by \(b\), there is no value in including \(a\) in the coalition. The ruler is already sure he will remain in power, and more importantly, \(a\)'s type is already known.

The ruler’s equilibrium expected value is constant for \(\omega \in [0, \bar{\omega}]\), falls discretely at \(\bar{\omega}\), and then increases linearly with beliefs. For \(\omega > \bar{\omega}\) the known-cost politician \(a\) has a strictly positive value \(V^*_a(1, \omega) = (1 - \omega)\delta \hat{b}_1 \phi\), because with probability \(1 - \omega\) \(b\) will be high-cost, in which case the separating offer will be rejected. The ruler will have to give rents to \(a\) with positive probability to remain in power.
5.2 One high-cost politician

The strategic environment both for the ruler and the politicians changes significantly when it is common knowledge that one of them is high-cost. Even if the ruler can credibly commit to make a pooling offer after a rejection by the known-cost politician, gaining his support requires a high offer of \( \overline{\theta} \), and this is only if the ruler is sure he would not be tempted to try to learn after a rejection. Facing a high-cost politician makes the commitment problem for the ruler worse, because the value of having the opponent politician reveal his type is higher. If after the known-cost politician has supported the ruler does not want to make an offer to the politician whose type is imperfectly known, then there is no learning just as in section 5.1. Now consider the alternative strategy in which the ruler makes first a pooling offer to the imperfectly-known-type politician. This offer is accepted with probability 1 and the ruler does not learn, just as when offering to the known-type politician first. And by leaving a known-cost politician to be offered second, the ruler is implicitly committing to make an offer which is accepted for sure in the event the imperfectly-known type of politician were to reject, so that no pivotality rents are required to achieve pooling by the politician whose type is not known in the first place. Now it is the politician with private information the one who “looks cheaper”. The analysis above implies that if after the acceptance of an offer by a known-to-be high-cost politician the ruler would not want to make an offer to the other politician (whose type is unknown), such a strategy can never do better for the ruler than the strategy of making first a pooling offer to the unknown-type politician.

**Proposition 7.** Assume \( \Omega \in \{ (\omega^a, \omega^b) : \omega^a \in (0, 1), \omega^b = 0 \} \). The MPBE can be characterized as follows: For \( \omega^a < \hat{\omega} \) the ruler makes a first offer \( m^1 = \overline{\theta} \) to \( b \) (the politician whose type is known to be high cost), which \( b \) accepts with probability 1 on the equilibrium path. Then the ruler makes a low-type-semi separating (LSS) second offer \( m^2 = \overline{\theta} + \frac{1}{2} \delta \phi \) to \( b \) which a high-cost \( a \) rejects and a low-cost \( a \) accepts with interior probability. Off-the-equilibrium path if \( b \) rejects, the ruler makes a pooling second offer \( m^2 = \overline{\theta} + \frac{1}{2} \delta \phi \) to \( a \) which both types of \( a \) accept. For \( \omega^a \geq \hat{\omega} \) the ruler makes a low-type semi separating first offer \( m^1 = \overline{\theta} \) to \( a \), which a high-cost type rejects and a low-cost type accepts with interior probability. After an acceptance decision is observed the ruler makes no second offer to \( b \), and if a rejection decision is observed the ruler makes a second offer \( m^2 = \overline{\theta} + \frac{1}{2} \delta \phi \) to \( b \), which \( b \) accepts. The cutoff value \( \hat{\omega} \) satisfies \( \hat{\omega} \in \frac{\delta (\mu - \overline{\theta}) - (1 - \delta) \overline{\theta} - (1 - \delta) \frac{1}{2} \phi}{\delta (\mu - \overline{\theta}) - (1 - \delta) \overline{\theta} - (1 - \delta) \frac{1}{2} \phi} \). Of course, the equilibrium is symmetric for \( \Omega \in \{ (\omega^a, \omega^b) : \omega^b \in (0, 1), \omega^a = 0 \} \).

**Proof.** See Appendix 3.
The dynamics under beliefs \((ω, 1)\) implied by Proposition 6 are quite different to those implied by Proposition 7 under beliefs \((ω, 0)\). Although along the equilibrium path the ruler remains in power under both kinds of beliefs, in the former case there is either no learning at all, or full learning in the first period (because the ruler makes a separating offer when \(ω\) is large). Under a state \((ω, 0)\), if the unknown-type politician is high-cost, beliefs will gradually fall, but only asymptotically will reach 0. If he is low-cost, beliefs might fall for some finite time but eventually the politician’s type will be revealed. The ruler will be “stubbornly” making a low offer to extract information, even for arbitrarily pessimistic beliefs (as long as the discount factor is large enough). Although the value of beliefs will not change the ruler’s willingness to test the unknown-type politician, it will determine the identity of the politician being made a first offer. Equilibrium supermajority coalitions can arise for exactly one period, for relatively pessimistic beliefs.

Figure 4 below depicts the dynamics for states \(Ω = (ω, 0)\) or \(Ω = (0, ω)\).

Figure 4: Equilibrium Dynamics at \(Ω = (0, ω)\) and \(Ω = (ω, 0)\) boundaries:

![Diagram](Image)

The intuition driving these dynamics highlights the trade-offs introduced by the strategic interaction between information extraction incentives and the threat of competition. With this purpose I will introduce some notation. Because the problem for the ruler is to choose a politician to make a first offer, and to chose the type of offer to the politician whose type is imperfectly known (the offer to a known-type politician is either 0 or such that he accepts it for sure), the strategy profiles can be written in the form \(γ = (t, Q)\), where \(t\) is the politician who is offered first, and \(Q \in \{PR, LSS, S, HSS, PA\}\) is the type of offer to the unknown-type politician. For illustrative purpose, assume \(a\) is the politician whose type is imperfectly known. Compare strategies \((a, Q)\) with strategies \((b, Q)\). Under \((a, Q)\) (when \(a\) is made a
first offer), no matter beliefs, a knows he cannot be pivotal, because if he rejects, the ruler will always be able to make an offer to b which is accepted, because b’s type is common knowledge. Also observe that if a is made the first offer, there will not be a supermajority coalition in equilibrium. This is because if a supports, the ruler will remain in power, and there is no incentive to make an offer to b given that his type is known. If a rejects, of course an offer to b will be made, but then the equilibrium coalition will only have b in it. The cost of making a the first offer, on the other hand, is that if a rejects it, then b becomes pivotal and the ruler will have to make him a large offer to dissuade him from triggering a power transition.

Compare this to a strategy \((b, Q)\) (where the first offer is made to b). After b has accepted, the ruler might still want to make an offer to a, hoping to make a reveal his type. This can be done cheaply because b has already accepted, so a is not pivotal. As a result, a supermajority coalition is possible. Of course, this has a cost because with positive probability the ruler will have given rents to two politicians despite requiring the support of only one to remain in power, but is potentially valuable if it achieves information extraction. On the other hand, a strategy which makes a first offer to b is also prone to creating a commitment problem for the ruler (which does not arise if a is offered first), for exactly the same reason the commitment problem arises when the known-type politician is low cost: If beliefs are optimistic enough, the ruler will be tempted, even off-path after a rejection by b, to make an informationally valuable offer to a. Knowing this, b will only be willing to support in the first place if given a pivotality rent.

The ruler has two possible no-learning strategies: \((b, PR)\) and \((a, PA)\) under which beliefs are absorbing\(^5\). In the former the ruler first makes an offer to b that b accepts, and then makes no offer to a on the equilibrium path. In the latter the ruler first makes a pooling offer to a that both types of a accept, and then makes no offer to b on the equilibrium path. Why the ruler might choose one over the other will depend on the behavior than can be sustained off-the-equilibrium path. Specifically, \((a, PA)\) is a “commitment” strategy. By leaving the known-cost type politician b to be offered last, the ruler is committing to making an offer which will be accepted for sure in case a rejects, no matter what beliefs might be. This makes a not-pivotal, and allows making the lowest possible pooling offer to a in the first place. The cost of following this strategy is, of course, that the ruler cannot learn a’s type.

As noted Appendix 3, under \((a, PA)\) the ruler would offer \(m = \bar{\theta}\) every period to a, which both types of a accept with probability 1. The value for the ruler is \(W_{(a, PA)}(\omega) = \frac{u - \bar{\theta}}{1 - \delta}\).

Just as under beliefs of the form \(\Omega = (1, \omega)\), if the ruler pursues the strategy \((b, PR)\) he will face a commitment problem for large values of \(\omega\). Off-the-equilibrium-path after a rejection by b, the ruler has incentives to make an offer that reveals some information about a’s type,

\(^5\)The strategy \((a, PR)\) is of course dominated by the strategy \((b, PR)\).
but which also implies the possibility of a power transition, increasing the ex-ante value of rejection for \(b\). This strategy is too risky for small values of \(\omega\), in which case the ruler prefers to make a pooling offer off-path. Under such range of beliefs strategies \((b, PR)\) and \((a, PA)\) are payoff-equivalent. For large enough \(\omega\), buying the support of \(b\) will require an offer larger than \(\tilde{\theta}\), so the ruler would do strictly better by making first a pooling offer to \(a\). The ruler can never do better by using a \((b, PR)\) strategy over a \((a, PA)\) strategy.

Why would the ruler ever want to follow a strategy under which the first offer is made to \(b\)? Assume that beliefs are low enough that off-path after a rejection by \(b\), the ruler will make a pooling offer to \(a\). Then, on the equilibrium path the ruler can try to learn \(a\)'s type cheaply, by making a low offer to \(a\), after \(b\) has supported \((a\) strategy of the form \((b, Q)\) with \(Q \neq PA\). Forging a supermajority coalition becomes an information extraction instrument. Commitment makes \(b\) not pivotal, and once \(b\) has accepted, \(a\) is not pivotal either, so both politicians will be willing to accept offers equal to their cost. When the ruler can commit to pool off the equilibrium path, this strategy can deliver a higher value for the ruler than a strategy in which he first makes an informationally valuable offer to \(a\), because an offer which has the potential to extract information must also imply a positive probability of rejection, which will make \(b\) pivotal on the equilibrium path. The ruler will then have to make a high offer to \(b\) to remain in power.

The proof of Proposition 7 (see Appendix 3) starts by showing that a strategy \((b, PA)\) cannot be part of an equilibrium: If it is optimal for the ruler to make a first offer to \(b\), it is never profitable to then induce pooling in support by \(a\). Because a first offer to \(b\) requires that the offer be accepted for sure, the ruler is already guaranteed to remain in power, while the subsequent offer to \(a\) is informationally worthless and requires a positive transfer.

Then it states that on the equilibrium path full separation is infeasible when the competing politician is known to be high-cost (for large enough \(\delta\)) (See Appendix 3). The reason is that under a separating equilibrium it is very attractive for a low-cost politician to reject and be believed to be high-cost, because from then on he will get an informational rent \(\tilde{\theta} - \tilde{\theta}\) half of the time. As a result, inducing separation requires making a very high offer, but such an offer makes the high-cost politician to also want to accept. The upwards incentive constraint cannot be satisfied. To see this, observe that a separating offer must satisfy the following incentive constraint for a low-cost type:

\[
m - \tilde{\theta} + \delta V_a^\alpha((1, 0)) \geq \delta V_b^\alpha((0, 0))
\]

which follows because if a low-cost \(a\) accepts, he reveals he is low cost, and beliefs are driven to \(\Omega = (1, 0)\), where his value is zero. If instead he rejects, in a separating equilibrium he will be believed to be high-cost, and beliefs are driven to \(\Omega = (0, 0)\). The low-cost politician
will then collect an informational rent half of the time, so \( V_{\bar{\theta}}((0,0)) = \frac{\bar{\theta} - \theta}{1 - \bar{\theta}} \). Thus, an offer that satisfies the low-cost incentive constraint must be at least \( m = \theta + \frac{\theta - \bar{\theta}}{1 - \bar{\theta}} \). Now look at the incentive constraint for a high-cost politician under a separating equilibrium:

\[
\delta V_{\bar{\theta}}((0,0)) \geq m^2 - \bar{\theta} + \delta V_{\bar{\theta}}((1,0))
\]  

which follows because a high-cost type should prefer to reject. The value for a high-cost politician is zero at \( \Omega = (0,0) \) and at \( \Omega = (1,0) \), so that he will be willing to reject the minimum offer that satisfies the low-cost incentive constraint iff \( \theta + \frac{\theta - \bar{\theta}}{1 - \bar{\theta}} - \bar{\theta} < 0 \), which requires \( \delta < 2/3 \). Thus, for a high \( \delta \) separation is infeasible. Notice that the infeasibility of separation in the current period is due to the very effective way in which the ruler can exploit competition in the future, once types have been revealed, effectively making the future under revelation to look bleak for a high-cost type.

Compare this situation to the one where the politician whose type is known happens to be low-cost. Separation was feasible then because a low-cost politician had little incentive to reject and be believed to be high cost, since he would be facing a low-cost politician and would get excluded from any coalition in the future. Being believed to be expensive is attractive only when the competing politician is expensive too, but in equilibrium this restricts the ability of the ruler to induce full separation. Conversely, the ruler has a harder time undermining the ratchet effect when a politician is known to be high cost.

The next step shows that a \textit{HSS} offer cannot be part of an equilibrium path. To show this I argue that transitioning from a \textit{LSS} strategy to a \textit{HSS} strategy is infeasible because the continuation value for a low-cost politician would be too high not to make him support with probability 1. Thus, \textit{HSS} offers can only follow \textit{HSS} offers. Then I show it is as costly to induce \textit{HSS} than to induce \textit{PS}, but under \textit{HSS} the ruler also faces the risk of having to give pivotality rents. The intuition is that an offer that induces \textit{HSS} is very costly to implement, and at the same time is not very attractive because it is likely to reveal negative information (after a rejection posterior beliefs must be \( \omega(R) = 0 \)).

Finally I show that the ruler prefers to induce \textit{LSS} over a no learning pooling strategy, even as beliefs become arbitrarily pessimistic (very likely that the politician is high-cost). The reason why as \( \omega \to 0 \) the ruler never prefers to stop testing \( a \) is that although by making a \textit{LSS} offer the ruler might incur the cost of a supermajority coalition, as \( \omega \) falls, the likelihood of actually having to make the transfer also falls as fast, so the expected payment falls as fast as the expected gain from achieving revelation. This predicts that even as a ruler becomes more and more pessimistic over time, he will stubbornly keep testing the politician with private information. To see this, observe that the ruler’s value of pursing a \textit{LSS} strategy is
given by

\[ W'(\Omega)|_{LSS} = \mu - \bar{\theta} + \omega \sigma^a [\delta W'(1, 0) - \theta] + (1 - \omega \sigma^a) \delta W'(\omega^a(R), 0) \]  

(11)

while \( \frac{\mu - \bar{\theta}}{1 - \delta} \) is the ruler’s value under a pooling offer. Equation (11) follows from the fact that with probability \( \omega \sigma^a \) the LSS offer will be accepted, revealing that \( a \) is low-cost, and driving posterior beliefs to \( \Omega = (1, 0) \), while with probability \( 1 - \omega \sigma^a \) the offer will be rejected, driving posterior beliefs about \( a \) down. While \( W'(1, 0) = \frac{\mu - \bar{\theta}}{1 - \delta} \), \( W'(\omega^a(R), 0) \geq \frac{\mu - \bar{\theta}}{1 - \delta} \) so that \( W'(\Omega)|_{LSS} > \frac{\mu - \bar{\theta}}{1 - \delta} \) for large enough \( \delta \), for any interior mixing strategy \( \sigma^a \).

The decision of making a first offer to \( b \) for a range of low values of \( \omega \), but a first offer to \( a \) for high values of \( \omega \), is driven by when the commitment problem becomes binding. For small \( \omega \) it is very risky to make a separating offer off the path because it is unlikely the politician will be low-cost, and the ruler might get ousted with high probability. Thus, when \( \omega \) is small the ruler can credibly threaten a pooling offer in case \( b \) rejects, making both politicians non-pivotal in equilibrium. For large values of \( \omega \) the ruler cannot commit not to make a separating offer after \( b \) has rejected, but he can instead make his first offer to \( a \), implicitly committing to remain in power, although at the cost of having to give pivotality rents to \( b \) with positive probability.

5.2.1 Constant Mixing Strategy

Recall from Proposition 7 and our discussion above, that equilibrium behavior at beliefs \( \Omega \in \{(\omega, 0), (0, \omega)\} \), prescribe the politician with private information to be made LSS offers of exactly \( m(\Omega) = \theta \) for any \( \omega \), where a low-cost type accepts with probability \( \sigma \), and a high-cost type always rejects. We did not have to specify the mixing probability of the low-cost type to characterize equilibrium behavior; any interior mixing probability is consistent with the ruler always wanting to induce LSS behavior. Interestingly, this leeway in specifying the speed with which a low type will reveal his type over time is a result of the nature of the strategic interaction: The ruler’s ability to exploit competition to undermine the ratchet effect incentives of the politician with private information implies that the value of both types of politician is driven down to zero. At the same time, for any interior mixing probability (which bayesian equilibrium requires to be correctly forecasted by all players), the competing politician will always strictly prefer to accept the equilibrium offer, so that the equilibrium strategy can be sustained for any probability of acceptance by the low-cost type.

The specific mixing probability by a low-cost politician will determine the cutoff value \( \hat{\omega} \), and the equilibrium value for the ruler. A constant mixing probability allows a straightforward analytic derivation of the players’ equilibrium values. Appendix 3 derives the equilibrium
cutoff value for $\sigma(\theta, \Omega) = 1/2$, which takes the following value:

\[ \hat{\omega} = \frac{\delta(\mu - \bar{\theta}) - (1 - \delta)\bar{\theta} - \delta(1 - \delta)\frac{1}{2}\phi}{\delta(\mu - \bar{\theta}) - (1 - \delta)\bar{\theta} - \delta(1 - \delta)\frac{1}{2}\phi - \delta \frac{\delta - \theta}{2}} \]  

(12)

Moreover, for $\omega < \hat{\omega}$, the ruler’s equilibrium value is given by

\[ W^r((0, \omega)) = \frac{\mu - \bar{\theta}}{1 - \delta} + \frac{\omega}{1 - \delta} \left[ \frac{\delta(\bar{\theta} - \theta)}{1 - \delta} - \theta \right] \]  

(13)

Equation (13) intuitively shows that when the ruler can credibly make a pooling offer following a rejection, his equilibrium value is increasing in his optimism about the politician with private information. The higher the beliefs, the more likely he will be able to avoid giving informational rents in the future.

5.3 Both Politicians’ costs imperfectly known

I now turn to the the characterization of equilibrium behavior for the more general case in which both politicians’ types are private information. A first key difference relative to when one of the politicians’ type is common knowledge, is that in such case the ruler can always make an offer that will be accepted for sure, precisely to the politician whose type is known. This allowed me to refer to strategy profiles simply with the identity of the first politician being offered, and the type of offer made to the imperfectly-known cost type politician (the offer to the known-cost type politician, and any off-the-equilibrium path offers, would follow trivially). When both politicians’ types are imperfectly known, the description of a strategy profile will require specifying the identity of the politician being offered first, the type of first offer, and the type of offers for the politician offered last, both after acceptance and rejection decisions are observed. Thus I will refer to strategy profiles in the form $\gamma = (\iota, Q_1, Q_A^1, Q_A^R)$, where the first element denotes the identity of the first politician being made an offer, the second element denotes the type of offer to politician $\iota$, and the third and fourth elements denote the type of offers to politician $\sim \iota$, after an acceptance and after a rejection decision by $\iota$, respectively. Again, $Q \in \{PR, LSS, S, HSS, PA\}$.

Proposition 8 below states the main result of the paper, and Figure 5 depicts it graphically.

**Proposition 8.** Assume $\Omega \in (0, 1) \times (0, 1)$. There exists a MPBE of the following form: For $\Omega \in \{(\omega^a, \omega^b) : \omega^a \geq \omega^b, \omega^b \leq \bar{\omega}\}$, $\gamma = (a, HSS, LSS, PA)$. For $\Omega \in \{(\omega^a, \omega^b) : \omega^a \geq \omega^b, \omega^b > \bar{\omega}, \omega^b \leq \hat{\omega}(\omega^a)\}$, $\gamma = (a, PS, PR, PS)$. For $\Omega \in \{(\omega^a, \omega^b) : \omega^a \geq \omega^b, \omega^b > \hat{\omega}(\omega^a)\}$, $\gamma = (b, S, S, PR)$. Symmetrically, for $\Omega \in \{(\omega^a, \omega^b) : \omega^a < \omega^b, \omega^a \leq \hat{\omega}\}$, $\gamma = (b, HSS, LSS, PA)$.
For $\Omega \in \{(\omega^a, \omega^b) : \omega^a > \omega^b, \omega^a \geq \omega^1(\omega^b)\}$, $\gamma = (b, PS, PR, PS)$. For $\Omega \in \{(\omega^a, \omega^b) : \omega^a < \omega^b, \omega^a \leq \omega^{-1}(\omega^b)\}$, $\gamma = (a, S, S, PR)$.

$\omega^i(\omega^a)$ is given by equation (14).

Proof. See Appendix 4.

Figure 5: Equilibrium Dynamics at $\Omega = (\omega^a, \omega^b)$

Key for the characterization of equilibrium for interior beliefs is the optimal behavior of the ruler at the interim stage after $i$'s action has been realized. Because full revelation takes beliefs to the boundary of the state-space, the detailed exposition of equilibrium behavior at the boundaries in section 4 and subsections 5.2 and 5.1 was necessary. This is because the cutoffs...
\( \bar{w} \) and \( \bar{\omega} \), which establish the limits at which the ruler starts facing a commitment problem, by leading to changes in equilibrium behavior at boundary beliefs, induce a partition of the state-space. Within different regions of the partition, different strategies can be sustained.

From the proof of Proposition 6 it follows that if \( \iota \) accepts an offer that reveals he is low cost at beliefs \( \Omega \), for \( \omega^{\sim \epsilon} \leq \bar{w} \) the ruler will prefer to make no second offer to \( \sim \iota \) at the interim stage. This follows directly from the fact that the ruler will remain in power for sure, and next period, under beliefs at \((1, \omega)\) or \((\omega, 1)\), he is fully entrenched and encroached. But moreover, Appendix 4 shows that for \( \omega^{\sim \epsilon} > \bar{w} \), the ruler will also make no second offer after an acceptance that reveals \( \iota \) is low cost. On the other hand, following a revealing rejection that drives posterior beliefs about the first politician being made an offer to 0, the ruler will find it optimal to make a fully separating offer (and thus risk being ousted) for sufficiently optimistic beliefs, while he will prefer to make a pooling offer if not optimistic enough. Recall that \( \bar{\omega} \) is precisely the value of beliefs at which the commitment constraint binds at the interim stage, and thus where the ruler switches his offering strategy (See Lemmas A4-2, A4-3 in Appendix 4).

Interim behavior at the boundary of the state-space is one key driver of equilibrium for interior beliefs. A second key element is the symmetric nature of Markov strategies, which imply that if at beliefs \((\omega, \omega')\) the ruler is following a strategy \((\iota, q, q^A, q^R)\), then at \((\omega', \omega)\) he must be following strategy \((\sim \iota, q, q^A, q^R)\). The symmetry of Markov strategies will put further restrictions on the equilibria that can be sustained at interior beliefs. In particular, the ruler will not be able to implement the type of strategy that can be sustained at beliefs of the form \( \Omega = (\omega, 0) \), where the known-to-be high-cost politician \( b \) was made a first offer that forced a low-cost \( a \) to accept and reveal his type with positive probability. In that case, no matter how low beliefs about politician \( b \) might fall, they would never be more pessimistic that beliefs about politician \( a \). In contrast, suppose that beliefs \((\omega^a, \omega^b)\) are interior and such that \( \omega^a > \omega^b \). If the ruler attempts a strategy of the form \((b, PA, \cdot, \cdot)\), in which he makes a first pooling offer to the politician under more pessimistic beliefs, and then makes an informative offer to the remaining politician, this will give incentives for a low-cost \( a \) to play a mixed strategy that drives posterior beliefs to be such that \( \omega^a(R) < \omega^b \). In this way, next period he will be the one being made a high uninformative offer and will collect informational rents (see Lemma A4-6 in Appendix 4). Notice that for posterior beliefs about \( a \) to fall enough as to become lower than beliefs about \( b \), the acceptance probability must be high enough, which reduces the value of \( a \) as it implies he will reveal his low type with a higher probability. Thus, this cannot be an equilibrium because for any given strategy \( \sigma^a \) such that \( \omega^a(R) < \omega^b \), \( a \) will always have a profitable deviation by choosing a slightly lower \( \sigma^a \) for which the posterior following a rejection is still slightly below \( \omega^b \). This was not a problem when \( \omega^b = 0 \), since in that case no matter the mixed strategy used by a low-cost \( a \), posterior
beliefs would always be higher than \( b \)'s. As a result, under relatively pessimistic beliefs about politician \( a \), the ruler will not be able to make \( LSS \) offers to \( a \). In equilibrium, \( HSS \) offers to the politician under more optimistic beliefs will be made first, which will allow the ruler to then make more informative \( LSS \) offers to the second politician. Thus, competition still has a partial capability of undermining the ratchet effect.

Highly optimistic beliefs about both politicians, on the other hand, imply that the ruler faces a commitment problem no matter which politician he makes a first offer to. Because politicians are sure to obtain pivotality rents in the future as a result, this allows the ruler to induce full separation of types in the present, because pivotality rents are the same for either type of politician. Equilibrium strategies will be of the form \((t, S, PR, S)\). Thus, learning is fast but implies the ruler will not be fully entrenched, since he will be ousted with positive probability. Lemma A4-7 shows that it is always optimal for the ruler in this range of beliefs to make a first offer to the politician under less optimistic beliefs. Finally, as beliefs become more pessimistic, the value of pursuing such a separating strategy falls quickly, because it is very likely a rejection will take place. \( \hat{\omega}(\omega^o) \) is a boundary condition that denotes the lowest beliefs about \( b \) for which the ruler is willing to pursue this strategy for every possible value of beliefs about \( a \). It will be implicitly determined by the following indifference condition (See Lemma A4-11):

\[
\omega^b = \frac{\delta \mu - \bar{\theta} - \omega^a [\delta (\mu - \theta) - (1 - \delta) \theta - \delta (1 - \delta) \frac{1}{2} \phi]}{(1 - \omega^a) [\delta (\mu - \theta) - (1 - \delta) \theta - \delta (1 - \delta) \frac{1}{2} \phi] + \omega^a \delta (1 - \delta) \frac{1}{2} \phi}
\]

(14)

The MPBE depicted above is of course, symmetric around the 45 degrees line. In the regions where the first offer is \( HSS \), beliefs are highly asymmetric, i.e., relatively pessimistic about one politician, and relatively optimistic about the competitor. The ruler makes a \( HSS \) first offer to the politician under more optimistic beliefs, which either reveals he is high cost, or shifts beliefs upwards. If an acceptance is observed, the ruler is then able to make a \( LSS \) offer to the remaining politician. Thus, the second politician either reveals he is low cost, or beliefs about him become more pessimistic. As long as neither politician fully reveals his type, beliefs asymptotically converge towards \((1, 0)\) or \((0, 1)\).

For relatively imprecise beliefs there is no learning. The ruler makes a first pooling offer \((PA)\) to the politician under more optimistic beliefs, and an offer to the competing politician only off path in case \( t \) rejects. This happens because for these intermediate beliefs, the rejection of a \( HSS \) offer drives beliefs to zero, at which point the ruler would actually prefer to make a separating offer to \( \sim t \). This commitment problem makes it very unattractive for the ruler to make an informative offer in the first place (recall that a first \( LSS \) offer is infeasible here). But beliefs are not optimistic enough to make the ruler willing to make a separating offer.
either.

Finally, for relatively optimistic beliefs, the ruler is unable to commit not to make a separating offer after a rejection, but he is optimistic enough that he prefers to make a separating offer in the first place, expecting full learning. As mentioned above, it is always optimal for the ruler in such case to make the first offer to the politician under more pessimistic beliefs. The reason is twofold. First, the expected cost of a rejection is higher at the stage in which the ruler can actually be ousted, so conditional on a rejection by \( t \) it is better for the ruler to be facing a politician \( \sim t \) who is more likely to accept a separating offer. Moreover, the separating offer to \( t \) must compensate him for forgoing the value of a power transition if he accepts. This expected value is decreasing in the opponent politician’s type, since the higher it is, the less likely a power transition takes place. As a result, the offer to \( t \) is decreasing in public beliefs about \( \sim t \). The ruler can be ousted with positive probability for sufficiently optimistic beliefs. Nevertheless, separation can be successful, so there is full learning along the equilibrium, at most in two periods. Figure 6 below depicts the learning dynamics under relatively pessimistic and relatively optimistic beliefs.

Figure 6: Learning under pessimistic and optimistic beliefs
6 Concluding Remarks

The increasing entrenchment and encroachment of rulers over time is an empirically widespread but theoretically puzzling phenomenon. This paper lays down a simple model to think about it, by studying the political-economy implications of the strategic interaction between political competition and ratchet effect incentives. I argued that in politically contested environments where coalitions must be forged and where providing political support is costly, incentives to learn about the costs of competing groups arise, since more precise information allows incumbents to capture a larger share of the spoils, and to hold a tighter grip on power. The paper thus investigated the extent to which rulers might be able to exploit the competitive nature of the coalition formation game to weaken the ratchet effect incentives of potential coalition members, as a function of public beliefs about the costliness of delivering political support by the competing groups. Thus, it intends to contribute to the understanding of the political economy of weakly institutionalized environments, and of the role of learning and informational asymmetries in coalition formation games.

The model delivers a rich set of predictions. First, it identifies a key trade-off between risk and learning, which yields an informational commitment problem that arises when rulers are relatively optimistic and hence face strong incentives to extract information. In the short-run, rulers’ eagerness to learn by making risky offers, even if only off-the-equilibrium path,
limits the power of competition to induce information revelation, because potential coalition members realize they might become pivotal with positive probability, and will only be willing to accept large, uninformative offers. In the long-run, on the other hand, the possibility of a future commitment problem actually allows a high degree of information revelation in the present. Potential coalition members are willing to accept low, informative offers today, because they will benefit from the commitment problem the ruler will face in the future. These backloaded incentives are powerful enough for revelation, when the ruler can exploit competition among potential coalition members.

The model also highlights that having certain information, even if negative, is a very powerful information extraction instrument in the short run, because for the ruler it becomes a source of commitment to remain in power, weakening the incentives of individual groups to attempt denning their support. It also implies that under some circumstances it will be attractive to include the more expensive, rather than the least expensive group, in the coalition, and that supermajority coalitions can emerge along the equilibrium path as information extraction mechanisms. In the long run, on the other hand, this will limit significantly the ability of a ruler to extract information through competition: If a group that is believed to be high cost will be included in future coalitions, low-cost groups will have a strong incentive to be believed high-cost by accepting only high offers in the present. But these high offers would be too attractive for low and high cost groups, becoming uninformative. As a result, the game is very discontinuous at the boundaries of the state-space, leading to radically different behavior when information about one group is common knowledge relative to when there is a slight lack of common knowledge.

Regarding the dynamic evolution of the game, the model shows that information will be revealed slowly under more pessimistic beliefs about these costs, but very fast under optimistic beliefs. The flip side of fast learning is that power transitions will be more likely, so we should expect power transitions to be more prevalent when rulers are optimistic and thus willing to face risks to learn. Moreover, learning negative information (from the ruler’s viewpoint) will be easier than learning positive information.

The model proposed in this paper is admittedly quite stylized. It would be of interest to allow for a larger number of potential coalition members, to study the effects of an increased amount of competition. By varying the majority requirements for remaining in power, this would also allow studying the effects of a heterogenous distribution of political power. The assumption of a discrete set of possible types was, of course, made for tractability, since introducing a continuum of types in a non-anonymous one-sided incomplete information game introduces several technical difficulties (see for example Schmidt (1993)). It remains an open question whether a richer type space would alter the results significantly.
Future empirical work should try to identify observable sources of variation in the cost of providing political support by different groups, and correlates of the precision of beliefs about these costs. These should allow to subsequently test predictions regarding the relationship between beliefs about the different politically relevant groups, and any observable patterns of distribution of resources and regime stability over time.
Appendix 1: The No-Competition Case

Equilibrium under no competition and complete information

Proposition 1. When $\omega = 0$ or $\omega = 1$, equilibrium strategies take the form

$$\sigma(m) = \begin{cases} 1 & \text{if } m \geq \theta + \phi(1 - \phi) \\ 0 & \text{otherwise} \end{cases} \quad m(1) = \theta + \phi(1 - \phi)$$

Proof. Let $\phi < \frac{\theta - \theta}{\phi(1 - \phi)}$, and $\omega = 1$. A necessary condition for politician $p$ to support the ruler with probability 1 is that the offer satisfies $V_{\theta}(1) = m - \theta + \delta V_{\theta}(1) \geq \delta \phi$.

In an equilibrium where $p$ accepts for sure an equilibrium offer $m$ at time $t$, he will accept the same offer every future $r > t$. The value for $p$ is thus $V_{\theta}(1) = \frac{m - \theta}{1 - \phi}$, and the equilibrium offer must satisfy $m \geq \theta + \phi(1 - \phi)$.

In equilibrium the ruler leaves no slack in $p$’s IC, so $m = \theta + \phi(1 - \phi)$ and the politician supports with probability 1. On the equilibrium path the ruler gets retained with probability 1, so his equilibrium value is

$$W^r(1) = \mu - m + \delta W^r(1) = \frac{\mu - m}{1 - \phi}$$

$$W^r(1) = \frac{\mu - \theta}{1 - \phi} - \delta \phi$$

Such an offer is sequentially rational iff the ruler prefers to make it than to grab all state resources and get ousted:

$$W^r(1) = \frac{\mu - \theta}{1 - \phi} - \delta \phi > \mu$$

which is equivalent to $\phi < \frac{\theta - \theta}{\phi(1 - \phi)}$. For $\omega = 0$, replace $\theta$ for $\theta$ everywhere above. Finally, the equilibrium value for the politician is $V_{\theta}(1) = \delta \phi$.

Equilibrium under no competition and incomplete information

Lemma 2: Strategies are decreasing in type: $\sigma(m, \omega) \leq \sigma(m, \omega)$.

Proof. Take a pair $(m, \omega)$ such that $\sigma(m, \omega) = 0$. Assume (by contradiction) that $\sigma(m, \omega) > 0$. Because a low-type politician is always rejecting, posterior beliefs after support has been observed must be that the politician is high-cost: $\omega(A) = 0$.

Because a low-cost politician rejects, it must be that $\phi > m - \theta + \delta V_{\theta}(0)$. Now observe that $V_{\theta}(0)$ is the value for a low cost politician when beliefs are that he is high cost for sure. From Appendix 1 we know that under this belief the ruler offers $m = \theta + \phi(1 - \phi)$. Thus,

$$V_{\theta}(0) = \frac{\theta + \phi(1 - \phi) - \mu}{1 - \phi} = \frac{\theta - \theta}{1 - \phi} + \delta \phi$$

Since a high-cost politician is supporting with positive probability, it must be that $m - \theta + \phi(1 - \phi) \geq \delta \phi$, which follows from the fact that after an acceptance, the posterior is $\omega = 0$. Together both ICs imply
\[ m - \bar{\theta} + \delta V_{\bar{\theta}}(0) \geq m - \theta + \delta V_{\theta}(0) \Leftrightarrow \delta^2 \phi - \delta \left( \frac{\bar{\theta} - \theta}{1 - \delta} + \delta \phi \right) \geq \bar{\theta} - \theta \Leftrightarrow \]

\[-\delta \theta - 1 - \delta \geq \bar{\theta} - \theta \Leftrightarrow -\delta \frac{1}{1 - \delta} \geq 0\]

which is a contradiction for any \( \delta \in (0, 1) \). Thus, if \( g(m, \omega) = 0, \bar{\sigma}(m, \omega) = 0 \) (a high-cost type cannot be accepting with positive probability if the low-cost type is rejecting for sure.)

Now take a pair \((m, \omega)\) such that \( g(m, \omega) \in (0, 1) \), and assume \( \bar{\sigma}(m, \omega) \in (0, 1) \). Both low-cost and high-cost politicians are mixing, so they must be indifferent between supporting and rejecting:

\[ V_{\bar{\theta}}(\omega) = m - \bar{\theta} + \delta V_{\bar{\theta}}(\omega(A)) = \delta \phi \]

and

\[ V_{\theta}(\omega) = m - \theta + \delta V_{\theta}(\omega(A)) = \delta \phi \]

Together these two imply

\[ m - \theta + \delta V_{\theta}(\omega(A)) = m - \bar{\theta} + \delta V_{\bar{\theta}}(\omega(A)) \Leftrightarrow \bar{\theta} - \theta = \delta V_{\bar{\theta}}(\omega(A)) - \delta V_{\theta}(\omega(A)) \]

which means that \( V_{\bar{\theta}}(\omega(A)) > V_{\theta}(\omega(A)) \). But notice that under beliefs \( \omega(A) \), whatever the offer \( m(\omega(A)) \) is, the low-cost politician can always mimic the high-cost politician’s strategy (the converse is not true). Thus, it must be that \( V_{\bar{\theta}}(\omega(A)) \geq V_{\theta}(\omega(A)) \), which yields a contradiction. Thus, when a low-cost politician is playing a mixed strategy, a high cost politician cannot be doing so too. He must be either supporting or rejecting with probability 1.

Thus, assume \( \bar{\sigma}(m, \omega) = 1 \). This means that posterior beliefs after support is observed must satisfy \( \omega(A) < \omega \). Because a low-cost politician is mixing he must be indifferent:

\[ V_{\bar{\theta}}(\omega) = m - \bar{\theta} + \delta V_{\bar{\theta}}(\omega(A)) = \delta \phi \]

Because a high-cost politician is supporting for sure, it must be that

\[ V_{\theta}(\omega) = m - \theta + \delta V_{\theta}(\omega(A)) \geq \delta \phi \]

Replacing the low cost’s indifference condition on the high-cost’s IC,

\[ m - \bar{\theta} + \delta V_{\bar{\theta}}(\omega(A)) \geq m - \theta + \delta V_{\theta}(\omega(A)) \]
\[
\delta [V_\bar{\theta}(\omega(A)) - V_\bar{\theta}(\omega(A))] \geq \bar{\theta} - \underline{\theta} \Leftrightarrow V_\bar{\theta}(\omega(A)) > V_\bar{\theta}(\omega(A))
\]

Which once more contradicts that \(V_\bar{\theta}(\omega(A)) \geq V_\bar{\theta}(\omega(A))\). We conclude that \(\sigma(m, \omega) \in (0, 1)\) implies \(\overline{\sigma}(m, \omega) = 0\). \(\square\)

**Lemma A1-1: Continuation equilibrium under offers** \(m \notin [0, \bar{\theta} + \delta(1 - \delta)\phi]\).

An equilibrium in which the ruler has made an offer \(m \notin [0, \bar{\theta} + \delta(1 - \delta)\phi]\) must be pooling: \(\forall \omega \in [0, 1]\),

\[
\sigma(m, \omega) = \begin{cases} 
0 & \text{if } m < \bar{\theta} \\
1 & \text{if } m \geq \bar{\theta} + \delta(1 - \delta)\phi
\end{cases} \quad \overline{\sigma}(m, \omega) = \begin{cases} 
0 & \text{if } m < \bar{\theta} \\
1 & \text{if } m \geq \bar{\theta} + \delta(1 - \delta)\phi
\end{cases}
\]

**Proof.** Neither type of politician can accept an offer below their cost. The ruler never makes an offer above \(\bar{\theta} + \delta(1 - \delta)\phi\), which is the offer he makes when he knows for sure the politician is high-cost, so a high-cost politician always accepts it. By pooling on support for offers above \(\bar{\theta} + \delta(1 - \delta)\phi\), a low-cost politician does not reveal his type. \(\square\)

Lemma 3 establishes that both types pool on acceptance when the offer is \(m \geq \bar{\theta} + \delta(1 - \delta)\phi\), and we know that high-cost types reject any offer below \(\bar{\theta}\), and low-cost types reject any offer below \(\bar{\theta}\). It thus remains to study the equilibrium behavior of a low-cost politician when facing an offer \(m \in [\bar{\theta}, \bar{\theta})\), and the equilibrium behavior of both types of politician when facing an offer \(m \in [\bar{\theta}, \bar{\theta} + \delta(1 - \delta)\phi]\).

**Lemma A1-2: Continuation equilibrium under offers** \(m \in [\bar{\theta}, \bar{\theta} + \delta(1 - \delta)\phi]\).

An equilibrium in which the ruler has offered \(m \in [\bar{\theta}, \bar{\theta} + \delta(1 - \delta)\phi]\) must be pooling on rejection:

\[
\sigma(m, \omega) = \overline{\sigma}(m, \omega) = 0
\]

**Proof.** Only continuation equilibria of the form 2) or 3) can exist, given that \(\overline{\sigma}(m, \omega) = 0\) (and the ruler will never make an offer which he knows will be reject for sure). Posterior beliefs after support is observed must be \(\omega(A) = 1\).

A necessary condition for such an equilibrium to exist is that the offer satisfies

\[
m - \bar{\theta} + \delta V_\phi(1) \geq \delta \phi
\]

Given \(V_\phi(1) = \delta \phi\), this condition is equivalent to \(m \geq \bar{\theta} + \delta(1 - \delta)\phi\). Thus, a low-cost politician will reject with probability 1. By Proposition 2 above, the high-cost type cannot be supporting with positive probability. \(\square\)
Lemma A1-3. Continuation equilibrium under offers $m \in [\theta + \delta(1 - \delta)\phi, \bar{\theta} + \delta(1 - \delta)\phi]$.

An equilibrium in which the ruler has offered $m \in [\theta + \delta(1 - \delta)\phi, \theta + \delta(1 - \delta)\phi]$ must be separating:

$$\sigma(m, \omega) = 1 \quad \bar{\sigma}(m, \omega) = 0$$

Proof. Separating Equilibrium:

The low-type’s IC is satisfied for any $m > \theta + \delta(1 - \delta)\phi$. By offering $m = \theta + \delta(1 - \delta)\phi$ the low-cost politician is willing to support.

If $\theta + \delta(1 - \delta)\phi < \bar{\theta}$, then the high-cost politician rejects for sure, and the ruler is retained with probability $\omega$. By deviating and offering any lower $m$ both types of politician reject and the ruler is ousted, so this cannot be a profitable deviation. By offering a higher $m$ the probability of remaining in power is the same, while the ruler is making a higher transfer, which cannot be profitable either. Thus, a separating equilibrium exists. To see that $m = \theta + \delta(1 - \delta)\phi$ is the unique equilibrium separating offer, consider any other larger $m'$; by deviating and offering $m = \theta + \delta(1 - \delta)\phi$ the ruler does not modify his probability of being retained, while he reduces the payment made, so this is a profitable deviation. Consider any offer $m'$ smaller than $m$. The ruler is ousted with probability 1, so it is always optimal to deviate and make an offer which is accepted for sure.

If $\theta + \delta(1 - \delta)\phi \geq \bar{\theta}$, the high-cost type is willing to reject if

$$\delta \phi > m - \bar{\theta} + \delta V_{\phi}(1)$$

$V_{\phi}(1) = \delta \phi$ because when the high-cost type is believed to be low cost he is offered $m = \theta + \delta(1 - \delta)\phi$, which, if he accepts, gives him a value of $\frac{\theta + \delta(1 - \delta)\phi - \bar{\theta}}{1 - \delta} < \delta \phi$, so that he always prefers to reject. Thus, for $m = \theta + \delta(1 - \delta)\phi$ the high-cost politician rejects with probability 1.

Low-cost Semi-Separating Equilibrium:

That such an equilibrium where $\sigma(m, \omega) \in (0, 1)$ does not exist can be shown as follows. Suppose in the continuation equilibrium the low-cost type politician is supporting with probability $\sigma(m, \omega)$. By Proposition 2, the high-cost type must be rejecting for sure. The following indifference condition must hold for a low-cost politician:

$$m - \theta + \delta V_{\phi}(1) = \delta \phi$$

which follows from the fact that after support the politician has revealed his type. Thus, the unique offer which can induce semi-separation is $m = \theta + \delta(1 - \delta)\phi$. Suppose that after such an offer the low-cost politician supports with probability $\sigma$. The value for the ruler is given by

$$W'(\omega) = \mu - \omega m + \delta \sigma W'(1) = \mu - \omega (\theta + \delta(1 - \delta)\phi) + \delta \sigma \left(\frac{\mu - \theta}{1 - \delta} - \delta \phi\right)$$

$$= \mu + \omega \sigma \left(\frac{\mu - \theta}{1 - \delta} - \frac{\theta - \delta^2 \phi}{1 - \delta}\right)$$

Now we can show that for any $\omega$ and $\sigma$, there always exists a positive $\varepsilon(\omega, \sigma)$ such that it is profitable for the ruler to deviate and offer an $m'$ such that $m < m' < \theta + \delta(1 - \delta)\phi + \varepsilon(\omega, \sigma)$. Under such an alternative offer the low-cost politician supports with probability 1. Thus, the value for the ruler would be

$$W'(\omega) = \mu - \omega m + \delta \sigma W'(1) = \mu - \omega (\theta + \delta(1 - \delta)\phi + \varepsilon(\omega, \sigma)) + \delta \sigma \left(\frac{\mu - \theta}{1 - \delta} - \delta \phi\right)$$
The deviation is profitable iff

\[ W^r_\epsilon^*(w) = \mu + \omega \left( \frac{\mu - \theta}{1 - \delta} - \delta^2 \phi - \epsilon(\omega, \sigma) \right) \]

Notice that \( \epsilon(\omega, \sigma) \) actually does not depend on \( \omega \), and that the term in square brackets is positive. Thus, for any \( \sigma > 0 \) the ruler can offer \( m' \) which yields a strictly higher payoff, and low-type semi-separation cannot be an equilibrium.

**High-cost Semi-separating equilibrium:**

That such an equilibrium where \( \sigma = 1 \) and \( \bar{\sigma}(m, \omega) \in (0, 1) \) does not exist can be shown as follows. Let \( m \geq \bar{\theta} \), so that both types can support. In such an equilibrium, posterior beliefs after support must be \( \omega(A) > \omega \). Because a high-cost politician is mixing, the following indifference condition must hold:

\[ m - \bar{\theta} + \delta V_\theta(\omega(A)) = \delta \phi \]

which implies the offer must be \( m = \bar{\theta} + \delta \phi - \delta V_\theta(\omega(A)) \). To proceed I make use of the following result:

**Lemma A1-4:** If at belief \( \omega \) the ruler makes an offer \( m \) which induces a High-cost semi-separating equilibrium, then the equilibrium cannot be Pooling on support for any \( \omega' > \omega \).

**Proof.** Let \( m(\omega) \) be an equilibrium offer under belief \( \omega \) which induces a High-cost semi-separating equilibrium, and suppose that at \( \omega(S) \) the ruler will make an equilibrium offer which induces pooling on support. The optimal such offer is \( m = \bar{\theta} + \delta(1 - \delta)\phi \), which both types accept for sure, and no learning obtains. Thus,

\[ W^r(\omega(A)) = \frac{\mu - \bar{\theta}}{1 - \delta} - \delta \phi \]

Now, because at \( \omega \) the equilibrium is semi-separating, the ruler’s value is

\[ W^r(\omega) = \mu + [\omega + (1 - \omega)\bar{\sigma}] (\delta W^r(\omega(S)) - m(\omega)) \]

Replacing the equilibrium value for the ruler under \( \omega(A) \), and using Bayes’ rule,

\[ W^r(\omega) = \mu + \frac{\omega}{\omega(A)} \left( \delta \frac{\mu - \bar{\theta}}{1 - \delta} - \delta \phi \right) - m(\omega) = \mu + \frac{\omega}{\omega(A)} \left( \delta \frac{\mu - \bar{\theta}}{1 - \delta} - \delta^2 \phi - m(\omega) \right) \]

Recall that the offer must be \( m(\omega) = \bar{\theta} + \delta \phi - \delta V_\theta(\omega(A)) \). Because at \( \omega(A) \) the ruler induces pooling, the value of a high-cost politician must be \( V_\theta(\omega(A)) = \delta \phi \). Replacing in the ruler’s value,
Because the ruler’s optimal offer is to induce semi-separation, it must be that \( W'(w) > \frac{\mu - \bar{\theta}}{1 - \delta} - \delta \phi \), which is the value of making an offer which both types accept for sure. Thus,

\[
\mu + \frac{\omega}{\omega(A)} = \mu + \frac{\omega}{\omega(A)} \left( \frac{\delta \mu - \bar{\theta} - \delta^2 \phi - (\bar{\theta} + \delta \phi - \delta^2 \phi)}{1 - \delta} \right) = \mu + \frac{\omega}{\omega(A)} \left( \frac{\delta \mu - \bar{\theta} - \bar{\theta} - \delta \phi}{1 - \delta} \right)
\]

Because that \( w(S) > w \), \( \omega(A) < 1 \), so the inequality above contradicts the optimality of semi-separation at \( w \).

The lemma above implies that if at beliefs \( \omega \) the ruler is making an offer that induces a high-cost semi-separating equilibrium, then at \( \omega(A) \) the equilibrium must be either high-cost semi-separating or separating. In either case \( V_\delta(\omega(A)) = \delta \phi \) because the high-cost must be indifferent between supporting and rejecting in the former case, and in the latter case he rejects for sure. The equilibrium offer for high-cost semi-separation must then be \( m = \bar{\theta} + \delta \phi - \delta^2 \phi = \bar{\theta} + \delta (1 - \delta) \phi \), which is exactly the offer that induces pooling in separation, and which a low-type accepts. Thus, no high-cost semi-separating equilibrium exists.
Appendix 2: Competition under Complete Information

Competition under complete information and symmetric types.

Proposition 4:

Assume $\Omega = (1, 1)$ (the derivation is identical for $\Omega = (0,0)$, except changing $\bar{\theta}$ for $\bar{\theta}$). Once the ruler has picked a politician ($p^1$) to make him a first offer, he faces the choice of whether to give $p^1$ an acceptable offer, or to make him an offer which he rejects for sure. The latter strategy cannot be ruled out a priori because the ruler is saving making a payment to $p^1$. This strategy implies nonetheless, making $p^2$ pivotal.

Assume the equilibrium is such that the ruler makes an offer $m^1 = 0$ to $p^1$, so that $\sigma^1 = 0$ in every period, and an offer $m^2$ to $p^2$ in the subgame after which $p^1$ has not supported, which $p^2$ accepts. The equilibrium value for $p^2$ is given by

$$V^2_\theta(\Omega) = m^2 - \theta + \delta \left[ \frac{1}{2} V^1_\theta(\Omega) + \frac{1}{2} V^2_\theta(\Omega) \right]$$

This expression takes into account that $p^2$ by accepting, rules out triggering a power transition. He gets the offer net of his cost, and next period with probability $1/2$ he is chosen as first to be offered, and with probability $1/2$ he is chosen as second to be offered. The equilibrium value for $p^1$ on the other hand, is given by

$$V^1_\theta(\Omega) = \delta \left[ \frac{1}{2} V^1_\theta(\Omega) + \frac{1}{2} V^2_\theta(\Omega) \right]$$

This expression comes from the fact that $p^1$ is given no offer in the current period, and in the next period he will be first to be offered with probability $1/2$, and second to be offered with probability $1/2$. This is equivalent to

$$V^1_\theta(\Omega) = \frac{\delta}{2 - \delta} V^2_\theta(\Omega)$$

Replacing this expression in the value for $p^2$,

$$V^2_\theta(\Omega) = m^2 - \theta + \delta \left[ \frac{\delta}{2 - \delta} V^2_\theta(\Omega) + \frac{1}{2} V^2_\theta(\Omega) \right]$$

$$V^2_\theta(\Omega) = m^2 - \theta + \frac{\delta}{2 - \delta} V^2_\theta(\Omega) = \frac{2 - \delta}{2(1 - \delta)} (m^2 - \theta)$$

The incentive constraint for $p^2$ requires that the equilibrium value is weakly larger than the value of the single-shot deviation in which $p^2$ rejects in the current period. Because $p^1$ did not support, a rejection by $p^2$ triggers a power transition for sure. The IC is:

$$V^2_\theta(\Omega) \geq \frac{1}{2} \sigma$$
The equilibrium value for the ruler is given by

\[ W'(\Omega) = \frac{\mu - m^2}{1 - \delta} \]

because every period the ruler offers \( m^2 \) to \( p^2 \), which he accepts. The IC for \( p^2 \) becomes

\[ \frac{2 - \delta}{2(1 - \delta)} (m^2 - \theta) \geq \frac{1}{2} \phi \iff \frac{2 - \delta}{1 - \delta} (m^2 - \theta) \geq \delta \phi \iff \]

\[ m^2 \geq \theta + \frac{\delta (1 - \delta) \phi}{2 - \delta} \]

Of course, the ruler makes the IC bind in equilibrium, so that \( m^2 = \theta + \frac{\delta (1 - \delta) \phi}{2 - \delta} \). In an equilibrium in which the ruler makes no offer to \( p^1 \) and pays for the support of \( p^2 \), his value is

\[ W'(\Omega) = \frac{\mu - (\theta + \frac{\delta (1 - \delta) \phi}{2 - \delta})}{1 - \delta} = \frac{\mu - \theta}{1 - \delta} - \frac{\delta \phi}{2 - \delta} \]

Now I look at an equilibrium where the ruler makes an acceptable offer to \( p^1 \). Assume the ruler made an offer \( m^1 \) to \( p^1 \) and that \( p^1 \) accepts the offer. The ruler now has to decide which offer, if any, to make to \( p^2 \). If the ruler makes an equilibrium offer \( m^2 \geq \theta \) which \( p^2 \) accepts, the value for \( p^2 \) is

\[ V^2(\Omega) = m^2 - \theta + \frac{1}{2} V^1(\Omega) + \frac{1}{2} V^2(\Omega) \]

This is because the ruler is retained with probability 1, and in the next period \( p^2 \) could be chosen as \( p^1 \) or \( p^2 \) with equal probabilities. The incentive constraint for \( p^2 \) to be willing to accept an offer \( m^2 \) is

\[ V^2(\Omega) \geq \delta \left[ \frac{1}{2} V^1(\Omega) + \frac{1}{2} V^2(\Omega) \right] \]

because, given that \( p^1 \) accepted, a rejection by \( p^2 \) does not produce a power transition. Replacing the equilibrium value for \( p^2 \),

\[ m^2 - \theta + \delta \left[ \frac{1}{2} V^1(\Omega) + \frac{1}{2} V^2(\Omega) \right] \geq \delta \left[ \frac{1}{2} V^1(\Omega) + \frac{1}{2} V^2(\Omega) \right] \]

\[ m^2 \geq \theta \]

This is, given that the continuation value for \( p^2 \) is invariant to his support decision, he is willing to accept any current offer which is at least as large as his cost. The lowest such offer is \( m^2 = \theta \). Now I argue that such an offer cannot be an equilibrium offer because the ruler has a profitable deviation. At the interim stage after \( p^1 \) has accepted, the value for the ruler is
\[ v'(Q, A) = \mu - m^1 - \theta + \delta W'(\Omega) \]

A profitable single-shot deviation for the ruler is to make no offer to \( p^2 \) and keep playing the equilibrium henceforth. This gives him a value of \( \mu - m^1 + \delta W'(\Omega) \), which is strictly larger than \( v'(Q, A) \). Thus, in an equilibrium where the ruler makes an acceptable offer to \( p^1 \), offering a positive transfer to \( p^2 \) cannot be part of the equilibrium after all; a supermajority coalition cannot be an equilibrium. In equilibrium it must be that \( m^2 = 0 \), and the value for \( p^2 \) is

\[ V^{2}(\Omega) = \frac{\mu}{2-\delta} V^{1}(\Omega). \]

Notice that after \( p^1 \) has accepted the ruler does not need the support of \( p^2 \) to remain in power. Nevertheless, the fact that he will not make positive offers to \( p^2 \) once \( p^1 \) has accepted makes \( p^1 \) pivotal in equilibrium, and suggests the equilibrium offer for \( p^1 \) to support should be high. This intuition is wrong because \( p^1 \) is not pivotal off the equilibrium path. If \( p^1 \) were to deviate and not support, the ruler will always find it optimal to credibly make an offer to \( p^2 \) which he accepts. This rules out a power transition and makes such a deviation by \( p^1 \) unprofitable. To see this, consider the stage where the ruler makes an offer to \( p^1 \), and let \( m^1 \) be an offer \( p^1 \) accepts. The equilibrium value for \( p^1 \) is

\[ V^{1}(\Omega) = m^1 - \theta + \delta \left[ \frac{1}{2} V^{1}(\Omega) + \frac{1}{2} V^{2}(\Omega) \right] \]

Replacing the equilibrium value for \( p^2 \),

\[ V^{2}(\Omega) = m^1 - \theta + \delta \left[ \frac{1}{2} V^{1}(\Omega) + \frac{\delta}{2-\delta} V^{2}(\Omega) \right] = \frac{2 - \delta}{2(1 - \delta)} (m^1 - \theta) \]

The incentive constraint for \( p^1 \) is simply

\[ V^{1}(\Omega) \geq \frac{\delta}{2 - \delta} V^{2}(\Omega) + \delta \frac{1}{2} V^{2}(\Omega) \]

because if \( p^1 \) rejects, the ruler will make an offer to \( p^2 \) which he will accept for sure (below I characterize such a credible off-the-equilibrium path offer). The IC reduces to

\[ V^{1}(\Omega) \geq \frac{\delta}{2 - \delta} V^{2}(\Omega) \]

Replacing the equilibrium value for \( p^2 \),

\[ V^{2}(\Omega) \geq \frac{\delta}{2 - \delta} \cdot \frac{\delta}{2 - \delta} V^{1}(\Omega) \iff 1 \geq \left[ \frac{\delta}{2 - \delta} \right]^2 \]

This inequality is true for any \( \delta \in [0, 1] \), so that any offer \( m^1 \geq \theta \) is accepted by \( p^1 \). The smallest such offer is \( m^1 = \theta \). The ruler can take \( p^1 \) down to his cost even though he is offering a Minimum Winning Coalition in equilibrium, because he has a credible threat of including \( p^2 \) in case \( p^1 \) rejects. The equilibrium value for the ruler of making an offer to \( p^1 \) is thus

\[ W'(\Omega) = \frac{\mu - \theta}{1 - \delta} \]

138
The ruler keeps all the surplus and is fully entrenched. Now, the ruler will prefer to make an offer to \( p^1 \) over making an offer only to \( p^2 \) iff

\[
\frac{\mu - \theta}{1 - \delta} > \frac{\mu - \theta}{1 - \delta} - \frac{\delta \phi}{2 - \delta}
\]

which holds for any \( \delta > 0 \). Thus, the unique sequentially rational equilibrium prescribes the ruler choosing a politician at random to make him a first offer which the politician accepts, and ignoring the second politician. The equilibrium values for the politicians are \( V_{p^1}(\Omega) = -\frac{2 - \delta}{\delta(1 - \delta)}(\theta - \bar{\theta}) = 0 \) and \( V_{p^2}(\Omega) = -\frac{\delta}{2 - \delta}V_{p^1}(\Omega) = 0. \)

Finally it remains to be checked that after a deviation in which \( p^1 \) rejects, the ruler credibly makes an offer to \( p^2 \) which \( p^2 \) is willing to accept. Assume \( p^1 \) has rejected. If the ruler makes an offer \( \hat{m}^2 \) to \( p^2 \) that he is willing to accept, such an offer must satisfy the following Incentive constraint for \( p^2 \):

\[
\hat{m}^2 - \theta + \delta \left[ \frac{1}{2} V_{p^1}(\Omega) + \frac{1}{2} V_{p^2}(\Omega) \right] \geq \delta \frac{1}{2} \phi
\]

which comes from the fact that after \( p^1 \)'s deviation and \( p^2 \)'s acceptance, the equilibrium keeps being played. If \( p^2 \) rejects, a power transition occurs, in which case he becomes the new ruler with probability \( 1/2 \), and otherwise he will be chosen as first to be offered with probability \( 1/2 \). The IC reduces to

\[
\hat{m}^2 \geq \bar{\theta} + \frac{\delta}{2} \phi
\]

The ruler leaves no slack in this constraint, so \( \hat{m}^2 = \bar{\theta} + \frac{\delta}{2} \phi \). Making this offer to \( p^2 \) is optimal for the ruler if his payoff is higher than taking all of state resources \( \mu \) and being ousted:

\[
\mu - \hat{m}^2 + \delta W'(\Omega) \geq \mu \Leftrightarrow \delta W'(\Omega) \geq \hat{m}^2 \Leftrightarrow
\]

\[
\frac{\delta(\mu - \bar{\theta})}{1 - \delta} \geq \bar{\theta} + \frac{\delta}{2} \phi \Leftrightarrow \delta(\mu - \bar{\theta}) \geq \bar{\theta} - \delta \bar{\theta} + \frac{1}{2} \delta(1 - \delta) \phi \Leftrightarrow \delta \mu \geq \bar{\theta} + \frac{1}{2} \delta(1 - \delta) \phi
\]

The equilibrium is identical for \( \Omega = (0, 0) \), just replacing \( \bar{\theta} \) for \( \theta \). Thus, the restriction for a credible threat by the ruler when \( \Omega = (0, 0) \) is \( \delta \mu \geq \bar{\theta} + \frac{1}{2} \delta(1 - \delta) \phi \), which implies the restriction above given that \( \bar{\theta} < \bar{\theta} \). Notice this inequality is implied by the original assumption \( \frac{\delta(\mu - \bar{\theta})}{\delta(1 - \delta)} > \phi \).

**Competition under complete information and asymmetric types.**

When one politician is high-cost and the other is low-cost, the optimal strategy for the ruler must optimally specify which politician to make an offer first. Without loss of generality assume \( \Omega = (0, 1) \), so that \( \theta^a = \bar{\theta} \) and \( \theta^b = \bar{\theta} \). I first prove the following lemma:
Lemma A2-1: Under $\Omega = (0, 1)$, if the ruler can credibly make an offer which is accepted with probability 1 to $p^2$ after $p^1$ has rejected, then $\ell = b$.

Proof. Assume the ruler can always make an offer to $p^2$ which $p^2$ accepts after $p^1$ has rejected. This implies that, at the stage game where $p^1$ must decided whether to support, rejecting can never give a higher payoff than accepting. This is because after a rejection $p^2$ will accept, so the continuation value is unchanged, whereas $p^1$ has forgone the current offer. Because the first politician who is offered an acceptable transfer (weakly higher than his cost) will always accept it, the ruler will always prefer to make a first offer to the low-cost politician ($b$ in this case).

Proposition 5:

Proof. The proof of proposition 4 showed that such a credible offer always exists, so that $\Upsilon((0, 1)) = 0$. The ruler offers $m^1 > \theta$ to $p^1$. If $b$ accepts, the ruler must decided what offer to make to $p^2 (= p^2)$. In an equilibrium where $p^2$ accepts the offer (which requires $m^2 \geq \theta$), his value is

$$V^2_{\theta}(\Omega) = \frac{m^2 - \theta}{1 - \delta}$$

The Incentive constraint for $p^2$ requires that not supporting is unprofitable:

$$V^2_{\theta}(\Omega) = \frac{m^2 - \theta}{1 - \delta} \geq \delta V^2_{\theta}(\Omega)$$

which is trivially true. This is because a rejection cannot trigger a power transition in a continuation game after which $p^1$ has accepted. The minimum such offer that makes $p^2$ accept with probability 1 is $m^2 = \bar{\theta}$. This gives the ruler an interim value of $v^2(\Omega, A) = \frac{\mu - m^1 - \bar{\theta}}{1 - \delta}$. Now observe that a deviation in which the ruler offers $m^2 = 0$ and otherwise the equilibrium is followed is profitable for the ruler at the interim stage because the ruler is already being retained:

$$v^2(\Omega, A) = \frac{\mu - m^1 - \bar{\theta}}{1 - \delta} < \mu - m^1 + \delta \frac{\mu - m^1 - \bar{\theta}}{1 - \delta}$$

Thus, making an offer to the second politician cannot be part of an equilibrium in which the ruler makes an acceptable offer to the first politician (the coalition must be minimum winning).

At the stage where the ruler makes an offer to $p^1$, let $m^1$ be an acceptable offer which is accepted with probability 1. On the equilibrium path the ruler remains in office, so the value for $p^1$ is

$$V^1_{\theta}(\Omega) = \frac{m^1 - \theta}{1 - \delta}$$

and his incentive constraint is

$$V^1_{\theta}(\Omega) = \frac{m^1 - \theta}{1 - \delta} \geq \delta V^2_{\theta}(\Omega)$$

which is trivially satisfied, and comes from the fact that if $p^1$ rejects, the ruler will make an offer to $p^2$ which $p^2$ will accept, so no political transition takes place and the game is identical from then on. The IC is satisfied for any $m^1$, and an acceptable offer requires $m^1 \geq \theta$. Thus, the ruler offers $m^1 = \bar{\theta}$, and $p^1$ supports.

It remains only to check that the ruler will want to offer $p^2$ a transfer which he accepts after $p^1$ has rejected. Let $\bar{m}^2$ be such offer. It must satisfy
\[ \bar{m}^2 - \bar{\theta} + \delta V^2_{\phi}(\Omega) \geq \delta \frac{1}{2} \phi \]

On the equilibrium \( p^2 \) is not included in any coalition, so \( V^2_{\phi}(\Omega) = 0 \). The IC for \( p^2 \) reduces to

\[ \bar{m}^2 \geq \bar{\theta} + \delta \frac{1}{2} \phi \]

The ruler leaves no slack on this constraint, setting \( \bar{m}^2 = \bar{\theta} + \frac{1}{2} \delta \phi \). This is optimal for the ruler as long as

\[ \mu - \bar{m}^2 + \delta W'_{\Phi}(\Omega) \geq \mu \Leftrightarrow \delta W'_{\Phi}(\Omega) \geq \bar{m}^2 \Leftrightarrow \]

\[ \frac{\delta \mu - \bar{\theta}}{1 - \delta} \geq \bar{\theta} + \frac{\delta}{2} \phi \Leftrightarrow \delta \mu \geq \bar{\theta} - \delta (\bar{\theta} - \bar{\theta}) + \frac{\delta}{2} (1 - \delta) \phi \]

Notice this inequality is implied by the original assumption \( \frac{\delta \mu - \bar{\theta}}{\delta (1 - \delta)} > \phi \).
Appendix 3: Competition under Incomplete Information

Entrenchment Dynamics at $\Omega = (1, \omega)$

Proposition 6:
Under $\Omega \in \{(1, \omega) : \omega \in (0, 1)\}$ the highest possible value the ruler could attain in any equilibrium is $W^* = \frac{\omega - \theta}{1 - \delta}$, where he gives one low-cost politician just his cost of providing political support, and this politician supports the ruler with probability 1 each period. If the ruler can sustain a strategy which achieves this value on the equilibrium path (call it the optimal strategy), he will pursue it.

I start proving the following:

Lemma A3-1: For any $\Omega \in \{ (\omega^a, \omega^b) : \omega^a = 1, \omega^b \in (0, 1) \}$, $V^a(\Omega) = 0$. (The equilibrium value of a high cost politician whose type is not perfectly known is zero when he faces a competitor whose type is known to be low-cost).

Proof. If it is optimal for the ruler to make an offer to $p^a$ first, given that $p^a$'s type is common knowledge, it must be that such offer is accepted with probability 1 by $p^a$ in equilibrium (the ruler knows which offer induces support, and if it is optimal not to induce support from $p^a$, the ruler will not make him any offer). The ruler finds it optimal to make no offer to $p^b$, so the equilibrium value of $b$ is zero no matter his type. If it is optimal for the ruler to make an offer to $p^b$ first, then after a rejection by $p^b$ the ruler can always make an offer that $p^a$ accepts with probability 1 because $p^a$'s type is known. No matter the action taken by $p^b$ the ruler remains in power, and the value of rejecting an offer is zero. Any offer which is at least as large as $p^a$'s type is thus accepted by $p^b$, so the ruler, if offering to $p^b$ first, will never offer more than $\tilde{\theta}$, which gives a high cost politician a value of zero.

Now consider the following strategy: $T(\Omega) = 1$ (make an offer to $p^a$, who is known to be low-cost, first), and offer him $m^1 = \bar{\theta}$. If $p^a$ supports, make no offer to $p^b$, and if $p^a$ rejects, make an offer to $p^b$ which induces pooling on support. Under this strategy, off-the-equilibrium-path after $p^a$ has rejected, the ruler is sure to achieve support of $p_b$, so that $V^a(\Omega) = 0$ (Claim A3-1). An offer $m^2 = +$ makes the high-cost $p^b$'s incentive constraint bind (and clearly makes the low-cost type's IC slack).

The interim value for the ruler of making this offer is given by

$$m^2 - \bar{\theta} + \delta V^a(\Omega) \geq \frac{1}{2} \phi$$

where I have made use of the fact that in a pooling equilibrium beliefs remain unchanged. Now, on-the-equilibrium path $p^a$ accepts the equilibrium offer after which the ruler makes no offer to $p^b$, so that $V^a(\Omega) = 0$ (Claim A3-1). An offer $m^2 = \bar{\theta} + \frac{1}{2} \phi$ makes the high-cost $p^b$'s incentive constraint bind (and clearly makes the low-cost type's IC slack).

The interim value for the ruler of making this offer is given by

142
\[ v'(Q, m^2) = \mu - (\bar{\theta} + \bar{\delta} \frac{1}{2} \phi) + \delta W'(\Omega) = \mu - \bar{\theta} - \delta \frac{1}{2} \phi + \delta \frac{\mu - \bar{\theta}}{1 - \delta} \]

where, once again, beliefs remain at \( \Omega \) after \( p^b \) supports.

**Lemma A3-2:** If at beliefs \((1, \omega)\) the ruler can sustain the optimal strategy, then he can sustain it for any \( \omega' < \omega \).

**Proof.** Suppose at \((1, \omega)\) the ruler can sustain the optimal strategy. This means that off-the-equilibrium-path after \( p^a \) has rejected, the ruler credibly induces pooling, so that for any other \( \bar{m}^2 \),

\[ v'(\Omega, \bar{m}^2) = \mu + (\delta W'(\Omega) - \bar{m}^2) \geq \mu + |\omega^b(\Omega, \bar{m}^2) + (1 - \omega)\delta^b(\Omega, \bar{m}^2)| \]

Observe that the pooling-equilibrium offer \( m^2 \) does not depend on \( \omega \). Thus, at \( \omega' \) the ruler can offer the same \( m^2 \) off-path, and induce pooling on support as well. Offering \( \bar{m}^2 > m^2 \) can never be optimal. Assume that at \( \omega' < \omega \) it is optimal for the ruler to offer \( \bar{m}^2 < m^2 \). If such \( \bar{m}^2 \) induces separation, then

\[ \mu + \omega'(\delta W'((1,1) - \bar{m}^2)) > \mu + (\delta W'(\Omega') - \bar{\theta} - \delta \frac{1}{2} \phi) = v'(\Omega, m^2) \]

No separation of \( p^b \) for \( \omega' < \omega \) (after a rejection by \( p^a \)). The binding incentive constraint for separation or low-cost semi-separation is \( m - \bar{\theta} + \delta V_b((1,1)) = \delta \frac{1}{2} \phi \), which also does not depend on \( \omega \). Thus, only \( \bar{m}^2 = \bar{\theta} + \delta \frac{1}{2} \phi \) is offered to induce such behavior by \( p^b \) for any \( \omega \).

The value for the ruler of inducing such behavior at \( \Omega \) is thus

\[ v'(\Omega, \bar{m}^2)|_{S} = \mu + \omega'(\delta W'((1,1) - \bar{m}^2)) > \mu + \omega'(\delta W'((1,1) - \bar{m}^2)) > v'(\Omega, m^2) \]

contradicting the optimality of inducing pooling at \( \Omega \).

No low-cost semi-separation of \( p^b \) for \( \omega' < \omega \) (after a rejection by \( p^a \)): The low-cost type politician must be indifferent:

\[ \bar{m}^2 - \bar{\theta} + \delta V_b((1,1)) = \delta \frac{1}{2} \phi \]

which requires \( \bar{m}^2 = \bar{\theta} + \delta \frac{1}{2} \phi \). Using the same argument as in Lemma A1-2, for any \( \omega \) there always exists some \( \epsilon > 0 \) such that the ruler could always offer \( \epsilon \) more and shift the support probability to \( 1 \), giving him a strictly higher value.

No high-cost semi-separation for \( \omega' < \omega \): The binding incentive constraint for high-cost type semi-separation at some \( \omega \) is \( m_{HSS}(\omega) - \bar{\theta} + V_b((1, \omega(A))) = \delta \frac{1}{2} \phi \). By claim A3-1 above, \( V_b((1, \omega(A))) = 0 \) for any \( \omega(A) \). Thus, \( m_{HSS} = \bar{\theta} + \delta \frac{1}{2} \phi = \bar{m}^2 \), which is the offer that induces pooling, and contradicts the fact that \( \bar{m}^2 < m^2 \).

Lemma A3-2 above implies that it will be sufficient to find the highest possible \( \omega \) for which pooling off-the-equilibrium-path is sustainable, to know the optimal behavior for any smaller \( \omega \).

The off-the-equilibrium path pooling-on-support strategy is sustainable only if the ruler has no profitable deviation. Consider a high-cost semi-separating deviation where \( \phi^b = 1 \) and \( \bar{\phi}^b \in (0, 1) \). In such a deviation posterior beliefs must be \( \omega(A) > \omega \), and the offer \( \bar{m}^2 \) must be such that the high-cost type is indifferent between support and rejection:

\[ \bar{m}^2 - \bar{\theta} + \delta V_b((1, \omega(A))) = \delta \frac{1}{2} \phi \]
From Lemma A3-1 above, we know that \( V_\delta((1, \omega(A))) = 0 \), which implies the offer \( m^2 = m^2 \) would induce pooling. Consider now a separating deviation where \( \sigma^b = 1 \) and \( \sigma^b = 0 \). Posterior beliefs would be \( \omega(A) = 1 \). Under such an offer the incentive constraint for a low-cost politician is

\[
m^2 - \theta + \delta V_\delta((1, 1)) \geq \frac{1}{2} \phi
\]

On the best possible deviation the ruler would leave no slack, and since \( V_\delta((1, 1)) = 0 \), \( m^2 = \theta + \frac{1}{2} \phi \), which a high-cost type cannot accept. The interim value for the ruler would be

\[
v'(\Omega|A) = \mu + \omega[\frac{\mu - \theta}{1 - \delta} - \theta - \frac{1}{2} \phi] = \mu + \omega[\frac{\mu - \theta}{1 - \delta} - \theta - \frac{1}{2} \phi] = v'(\Omega, m^2) \Leftrightarrow
\]

\[
\omega > \frac{\delta(\mu - \theta) - (1 - \delta)\bar{\theta} - \delta(1 - \delta)\frac{1}{2} \phi}{\delta(\mu - \theta) - (1 - \delta)\bar{\theta} - \delta(1 - \delta)\frac{1}{2} \phi} \equiv \bar{\omega}(\delta) \in (0, 1)
\]

We conclude that for beliefs \( \omega \leq \bar{\omega}(\delta) \) the unique MPBE prescribes the optimal strategy.

Let \( \omega > \bar{\omega}(\delta) \). Only two types of strategy profiles for the ruler can be an equilibrium. Either the ruler offers to \( p^a \) first (strategy a)), in which case he must make a high enough offer that \( p^a \) will accept with probability 1, and off-the-equilibrium path he separates, or he offers first to \( p^b \) (strategy b)), and only second to \( p^a \) in case of a rejection by \( p^b \).

Consider strategy a) in which \( T(\Omega) = 1 \), and the ruler offers \( m^2 \) such that \( p^a \) accepts. From the previous analysis we know that in case of a rejection, the ruler will offer a separating offer \( m^2 = \theta + \frac{1}{2} \phi \). Since \( \delta(1 - \delta) \frac{1}{2} \phi \) for sure, and a high-type rejects. Posterior beliefs after support are \( \omega(A) = 1 \). Thus, off-the-equilibrium path with probability \( 1 - \omega \) the ruler is ousted. This implies that \( p^a \)'s incentive constraint is given by

\[
m^1 - \theta + \delta V_\delta((1, 1)) \geq \omega \delta V_\delta((1, 1)) + (1 - \omega)\delta \frac{1}{2} \phi
\]

This IC comes from the fact that if \( p^a \) accepts, he gets the transfer net of his cost, and the ruler makes no offer to \( p^b \) so that beliefs remain unchanged. If he were to reject, the ruler would make a separating offer to \( p^b \), which will be accepted if he is low-cost (with probability \( \omega \)), in which case posterior beliefs will be \( \Omega = (1, 1) \), and will be rejected if he is high-cost (with probability \( 1 - \omega \)), in which case the ruler is ousted. In equilibrium \( p^a \) is accepting in every period, so the binding incentive constraint can be expressed as

\[
V_\delta(\Omega) = \frac{m^1 - \theta}{1 - \delta} = (1 - \omega)\delta \frac{1}{2} \phi \Leftrightarrow m^1 = \theta + (1 - \omega)(1 - \delta)\delta \frac{1}{2} \phi
\]

The offer which makes \( p^b \) be willing to support is decreasing in \( \omega \), because as beliefs increase the probability of a power transition falls after \( p^a \) has rejected. The equilibrium value for the ruler is
Notice that on the equilibrium path there would be no learning. The value for the ruler increases as $\omega \to 1$. The description of equilibrium strategies also requires off-the-equilibrium path prescriptions. If $a$ were offered second after support, he would not be pivotal and would accept an offer $m = \theta$. If he were offered second after a rejection, he would be pivotal and would accept an offer that satisfies the interim IC $m = \theta + \delta V_a(\Omega) \geq \frac{1}{2} \phi$, where I have assumed off-path posterior beliefs after $p^b$ rejected are the same as the prior. The lowest $m$ satisfying this IC is $m = \theta + \delta [1 - \delta(1 - \omega)] \frac{1}{2} \phi$.

Off path, if $b$ is offered first, he knows $a$ will support if he accepts, so any type of $p^b$ is willing to support if given their cost. The alternative strategy $b$ makes the ruler exploit competition between politicians to extract information. By making an offer first to the politician whose type is not perfectly known ($p^b$ in this case), the ruler is indirectly committing to remain in power because if $p^b$ rejects, the ruler is left with making an offer to a politician whose type he perfectly knows, and thus, can always make an offer which is accepted by him. This reduces the outside value of both types of $p^b$ to zero because no power transition can take place. So consider a strategy with $T(\Omega) = 0$. On or off-the-equilibrium path after $p^b$ has rejected and interim beliefs are $(1, \omega(R))$, the ruler faces a low-cost politician who is willing to accept any offer satisfying

$$m = \theta + \delta V_a(1, \omega(R)) \geq \frac{1}{2} \phi$$

An offer $m = \theta - \delta V_a(1, \omega(R)) + \frac{1}{2} \phi$ is enough for $p^b$ to support. Notice that this offer is bounded above by $\theta + \delta \frac{1}{2} \phi$, which we know is an offer the ruler always wants to make to remain in power rather than being ousted for sure. Thus, the ruler will always make this offer in a subgame after which $p^b$ rejected.

Assume the equilibrium offer to $p^b$ induces pooling by $p^b$. The offer must satisfy the high-cost type incentive constraint:

$$m^1 - \theta + \delta V_a(1, \omega(R)) \geq \delta V_a(1, \omega(R))$$

Here it is necessary to specify off-the-equilibrium-path beliefs about $p^b$ after a rejection, given that the equilibrium prescribes support by both low and high cost politicians. If we assume $\omega(R) = 0$ or $\omega(R) = 1$, $V_a(1, \omega(R)) = 0$ so that an offer $m^1 = \theta$ would be enough to induce pooling by $p^b$. By assuming $\omega(R) = \omega$, $m^1 = \theta$ also satisfies the high-cost type’s IC (assumptions about off path beliefs are irrelevant). In this proposed equilibrium the ruler would offer $m^1 = \theta$ to $p^b$ every period, which both types of $p^b$ would accept with probability 1 on the equilibrium path. The value for the ruler of pursuing this strategy would then be $W(\Omega)|_{b, p} = \mu - \theta$. Pooling under $b$ is preferred to $a$ iff, for some $\omega > \bar{\omega},$

$$W(\Omega)|_{b, p} = \mu - \theta > \mu - \theta - (1 - \omega) \delta \frac{1}{2} \phi = W(\Omega)|_{a}$$

This inequality is most likely to hold for the smallest possible $\omega$, which is $\bar{\omega}$. Thus, if it does not hold for $\bar{\omega}$ it will not hold for any $\omega$. Replacing for $\bar{\omega}$ and rearranging,

$$\mu - \bar{\theta} > \mu - \theta - (1 - \omega) \delta \frac{1}{2} \phi \iff \frac{(1 - \delta)(\bar{\theta} - \theta)}{\delta(\mu - \theta) - (1 - \delta) \theta - \delta(1 - \delta) \frac{1}{2} \phi} > \frac{1 - \delta}{\delta(\mu - \theta) - (1 - \delta) \theta - \delta(1 - \delta) \frac{1}{2} \phi} \iff$$

$$\delta(1 - \delta) \frac{1}{2} \phi > \delta(\mu - \theta) - (1 - \delta) \theta - \delta(1 - \delta) \frac{1}{2} \phi \iff (1 - \delta)(\bar{\theta} - \theta) > \frac{1 - \delta}{\delta(\mu - \theta) - (1 - \delta) \theta - \delta(1 - \delta) \frac{1}{2} \phi} \iff$$

$$\delta(1 - \delta)^2 \phi > \delta(\mu - \theta) - (1 - \delta) \theta - \delta(1 - \delta)^2 \phi \iff (2 - \delta) \delta(1 - \delta) \frac{1}{2} \phi > \delta \mu - \theta$$
\[
\frac{\delta\mu - \bar{\theta}}{\delta(1 - \delta)} < (2 - \delta)\frac{1}{2}\phi
\]

Notice that \((2 - \delta)\frac{1}{2} < 1\). Thus, the inequality above does not hold given that we have assumed \(\frac{\delta\mu - \bar{\theta}}{\delta(1 - \delta)} > \phi\). Making a pooling offer to \(p^h\) first cannot be part of an equilibrium for any \(\omega\).

Assume the equilibrium offer to \(p^h\) induces separation. The offer must satisfy the low-cost type incentive constraint:

\[
m^1 - \bar{\theta} + \delta V^h_\theta((1, 1)) \geq \delta V^h_\theta((1, 0))
\]

which comes from the fact that in a separating equilibrium, \(\omega(A) = 1\) and \(\omega(R) = 0\). After acceptance the ruler makes no offer to \(p^a\), and after a rejection the ruler makes an offer to \(p^a\) which he accepts for sure. \(V^a_\theta((1, 0)) = 0\) because at \(\Omega = (1, 0)\) the ruler never includes \(b\) in a coalition. Thus, the offer \(m^1 = \bar{\theta}\) is enough to induce separation (this offer trivially satisfies the IC for not supporting by a high-cost type). The value for the ruler of following this strategy is:

\[
W^r(\Omega)|_{\text{LSS}} = \mu + \omega |\delta W^r'((1, 1)) - \bar{\theta}| + (1 - \omega) \left[\delta W^r'((1, 0)) - \bar{\theta} - \delta\frac{1}{2}\phi\right] = \mu - \bar{\theta} + \delta W^r' - (1 - \omega)\delta\frac{1}{2}\phi = W^r(\Omega)|_{\text{LSS}}
\]

Surprisingly, the ruler’s value of pursuing this strategy is exactly the same as the value of strategy a) for any \(\omega\).

Assume the equilibrium offer to \(p^h\) induces low-cost semi-separation, where \(\sigma(\Omega, m) \in (0, 1)\) and \(\bar{\sigma}(\Omega, m) = 0\). Posterior beliefs are \(\omega(A) = 1\) and \(\omega(R) < \omega\). The IC for the low-cost type is

\[
m^1 - \bar{\theta} + \delta V^h_\theta((1, 1)) \geq \delta V^h_\theta((1, \omega(R)))
\]

The high-cost type’s IC is

\[
\delta V^h_\theta((1, \omega(R))) \geq m^1 - \bar{\theta} + \delta V^h_\theta((1, 1))
\]

\(m^1 = \bar{\theta} + \delta V^h_\theta((1, \omega(R)))\) leaves no slack on the low-cost’s IC, and relaxes the high-cost’s IC as much as possible. Replacing in the high-cost’s IC, a necessary condition for a LSS equilibrium to exist is

\[
\bar{\theta} - \bar{\theta} \geq \delta \left[V^h_\theta((1, \omega(R))) - V^h_\theta((1, \omega(R)))\right]
\]

Notice that the RHS of this inequality is positive (because for any beliefs the value of the low-cost type cannot be smaller than the value for the high-cost type, given that the former could always mimic the latter). Thus, if the difference between values for low and high violates this inequality (recall from Claim A3-1 that the value for a high-cost type is zero for any \(\Omega = (1, \omega)\)), there exists no semi-separating equilibrium. Suppose the inequality holds. The ruler’s value of pursuing this strategy is

\[
W^r(\Omega)|_{\text{LSS}} = \mu + \omega \sigma(\Omega, m^1) \left[\delta W^r'((1, 1)) - \bar{\theta} - \delta V^h_\theta((1, \omega(R)))\right]
\]

\[
+ (1 - \omega \sigma(\Omega, m^1)) \left[\delta W^r'((1, \omega(R))) - \bar{\theta} + \delta V^h_\theta((1, \omega(R))) - \delta\frac{1}{2}\phi\right]
\]

which follows because with probability \(\omega \sigma(\Omega, m^1)\) \(p^h\) is low cost and accepts the offer, and with complementary probability \(p^h\) rejects the offer, in which case the ruler makes an offer to \(p^a\), who accepts for sure. This takes into
account that at such subgame beliefs have evolved to \( \Omega = (1, \omega(R)) \). Now I show that inducing separation is always preferred by the ruler to inducing semi-separation, by comparing \( W^r(\Omega)_{b|b} \) to \( W^r(\Omega)_{b|LSS} \). First, observe that net of \( \mu, W^r(\Omega)_{b|b} \) is the convex combination of the two terms in square brackets, with weights \( \omega \) and \( 1 - \omega \). Now take \( W^r(\Omega)_{b|LSS} \). Net of \( \mu \) it is also a convex combination, this time with weights \( \omega \) and \( 1 - \omega \). Clearly \( \delta W^r((1,1)) - \theta > \delta W^r((1,1)) - \delta V^f_{\theta}((1, \omega(R))) \) for any positive \( V^f_{\theta}((1, \omega(R))) \). This follows because at \( \Omega = (1,0) \) the equilibrium is efficient: the ruler is entrenched with probability 1 and is leaving no rents to either politician. On the other hand, if under \( \omega(R) \) the ruler remains in power with probability 1, then social surplus is the same as under \( \Omega = (1,0) \), and the sum of payoffs for ruler and \( p^b \) cannot exceed it. If under \( \omega(R) \) the ruler is ousted with positive probability, then total surplus is lower than under \( \Omega = (1,0) \). Thus, the term multiplying \( 1 - \omega \) is larger than the term multiplying \( 1 - \omega \). This means that \( W^r(\Omega)_{b|LSS} \) is the convex combination of two terms, each of which is smaller than the terms in the convex combination for \( W^r(\Omega)_{b|b} \). Now, in both \( W^r(\Omega)_{b|b} \) and \( W^r(\Omega)_{b|LSS} \) the first term in brackets is larger than the second term in brackets. Since \( \omega > \omega \) for any \( \omega \in (0,1) \), \( W^r(\Omega)_{b|LSS} \) puts less weight in the larger term, and more weight in the smaller term than \( W^r(\Omega)_{b|b} \). Thus, \( W^r(\Omega)_{b|LSS} > W^r(\Omega)_{b|b} \) for any \( \Omega = (1,\omega) \), so that low-cost semi-separation is not an equilibrium for \( \omega > \omega \).

Now take a high-cost semi-separating candidate equilibrium. Under such an equilibrium \( \sigma(\omega, m^4) = 1 \) and \( \sigma(\omega, m^4) \in (0,1) \). Posterior beliefs are \( \omega(A) > \omega \) and \( \omega(R) = 0 \). The incentive constraint for a low-cost \( p^b \) is

\[
m^1 - \theta + \delta V^f_{\theta}((1, \omega(A))) \geq \delta V^f_{\theta}(1,0))
\]

The indifference condition for a high-cost politician is

\[
m^1 - \tilde{\theta} + \delta V^f_{\theta}((1, \omega(A))) = \delta V^f_{\theta}(1,0))
\]

\( V^f_{\theta}(1,0) = V^f_{\theta}(1,0) = 0 \) because if \( p^b \) is believed to be high-cost he is never included in a supporting coalition. From Claim A3-1 we also have that \( V^f_{\theta}((1, \omega(A))) = 0 \), so that the high-cost semi-separating offer must be \( m^1 = \tilde{\theta} \). But this is the offer which induces pooling which, as shown above is never optimal.

We conclude that if the ruler makes an offer first to the politician whose type is not perfectly known, it must be a separating offer. Moreover, that strategy yields the same value for the ruler as the strategy in which he offers first to the politician whose type is known to be low-cost; strategies a) and b) are payoff equivalent, and in both the ruler remains in power with probability 1.

**Entrenchment Dynamics at \( \Omega = (\omega,0) \).**

**Proposition 7:**

Start by observing that \( W^r((0,0)) = \frac{\omega}{1-\omega} \) while \( W^r((1,0)) = \frac{\omega}{1-\omega} \). The ruler clearly prefers the state in which he faces one low and one high cost politicians to the state where he faces two high-cost politicians. For any \( \omega \in (0,1) \), \( p^b \) has private information about his cost, while it is common knowledge that \( p^b \) is high-cost. To characterize equilibrium dynamics I start by taking the set of feasible strategies for the ruler, and restrict it using the equilibrium conditions implied by the model. Recall a Markov strategy for the ruler prescribes choosing a politician \( t \in \{a, b\} \) to make a first offer, the value of the first offer \( m^1 \), and the value of the second offer \( m^2 \) to the remaining politician. Moreover, we know that the politician’s equilibrium strategies must be pooling on rejection (PR) where both low and high-cost types reject, low-cost semi-separating (LSS) where the low-cost mixes and the high-cost rejects, separating (S) where the low-cost supports and the high cost rejects, high-cost semi-separating (HSS) where the low-cost supports and the high-cost mixes, or pooling on support (PA) where both low and high-cost types support. Given that \( p^b \)'s type
is known, an equilibrium offer to \( p^b \) must be such that he accepts. As a result we can denote by \((t, Q)\) the strategy type used by the ruler, where \( Q \in \{PR, LSS, S, HSS, PA\}\).

**Lemma A3-3:** In equilibrium the ruler never uses a strategy \((b, PA)\).

*Proof.* Assume that in state \( \Omega = (\omega, 0) \) the ruler uses a strategy \((b, PA)\). Under such a strategy the ruler first makes an offer \( m^1 \) to \( p^b \), which must be accepted, and then makes an offer \( m^2 \) to \( p^a \) which is accepted by both low and high-cost types and hence leaves beliefs unchanged. Because beliefs \((\omega, 0)\) are absorbing, the ruler uses the same strategy \((b, PA)\) and makes the same offer in every future period. The value for the ruler is \( W_{(b,PA)}^r(\Omega) = \frac{\mu - m^1 - m^2}{1- \delta} \).

Now consider the strategy \((b, PR)\). Since the support decision by \( p^b \) comes before the ruler makes an offer to \( p^a \), \( m^1 \) must also be sufficient for \( p^b \) to support when the ruler makes a \( PR \) offer to \( p^a \), for whatever the off-the-equilibrium path behavior after a rejection is. Beliefs about \( p^a \) also remain unchanged, while the ruler is not making a payment to \( p^a \). The value of such a strategy is \( W_{(b,PR)}^r(\Omega) = \frac{\mu - m^1}{1- \delta} \) which is strictly larger than \( W_{(b,PA)}^r(\Omega) \).

\[\square\]

**Lemma A3-4:** The ruler can never induce full separation on the equilibrium path: No strategy \((t, S)\) can be part of an equilibrium for any \( \omega \).

*Proof.* A necessary condition for a strategy \((t, S)\) to be part of an equilibrium is that a low-cost type wants to support:

\[ m - \theta + \delta V^a_p((1,0)) \geq \delta V^a_p((0,0)) \]

which follows because if a low-cost type \( p^a \) supports in a separating equilibrium, he reveals his type, so posterior beliefs are \( \omega(A) = 1 \). If he rejects, because \( p^b \) has already accepted (if \( t = b \)) or will accept (if \( t = a \)), there is no power transition and posterior beliefs are that he is high-cost \( \omega(R) = 0 \). Under \( \Omega = (0,0) \) the ruler offers \( m^1 = \bar{\theta} \) (see Proposition 4) to a politician chosen at random. This means that a low-cost type’s value of being believed high-cost is \( V^a_p((0,0)) = \frac{\frac{\theta}{2}}{1- \delta} \). Recall also that under \( \Omega = (1,0) \) the ruler offers \( m^1 = \theta \) to \( p^a \), so that \( V^a_p((1,0)) = 0 \). Replacing these conditions in the IC above,

\[ m^2 \geq \theta + \delta \frac{\bar{\theta} - \theta}{2(1- \delta)} \]

Another necessary condition for separation is that a high-cost type wants to reject:

\[ \delta V^a_p((0,0)) \geq m^2 - \bar{\theta} + \delta V^a_p((1,0)) \]

Under \( \Omega = (0,0) \) the equilibrium value for \( \bar{\theta} \) is \( V^a_p((0,0)) = 0 \), and notice that \( V^a_p((1,0)) = 0 \) too. The latter follows because under \( \Omega = (1,0) \) the ruler offers \( m^1 = \bar{\theta} \) to \( p^a \), but this offer cannot be accepted by a high-cost \( p^a \), given that \( \theta < \bar{\theta} \). \( p^a \) must reject, after which the ruler makes an off-the-equilibrium-path offer to \( p^b \) which \( p^b \) accepts, and this happens in every future period. The high-cost type’s IC is simply \( 0 \geq m^2 - \bar{\theta} \). Reducing \( m^2 \) until the low-cost type’s IC binds makes the high-cost type’s IC most likely to hold. For \( m^2 = \theta + \delta \frac{\bar{\theta} - \theta}{2(1- \delta)} \),

\[ 0 \geq \theta + \delta \frac{\bar{\theta} - \theta}{2(1- \delta)} - \bar{\theta} \iff 0 \geq \frac{\delta}{2(1- \delta)} - 1(\bar{\theta} - \theta) \iff \]

\[ 0 \geq \frac{\delta}{2(1- \delta)} - 2(1- \delta) \iff 0 \geq 3\delta - 2 \iff \delta \leq \frac{2}{3} \]

148
For any $\delta$ larger than $\frac{2}{3}$ separation cannot be sustained.

Since we are considering $\delta \to 1$, separation cannot be an equilibrium. The reason is that to induce separation the ruler must make an offer high enough that a low-type is willing to reveal his type. This is costly because if the low-type were believed high-cost he would get an informational rent every period, whereas after revelation his value is driven to zero. But for such a high offer, the high-cost type’s IC cannot be satisfied; he also prefers to support for large enough discount factors. This is analogous to the standard Laffont and Tirole (1988) “too much pooling” result. Notice this argument follows no matter if the ruler’s first offer is to $a$ or to $b$.

**Lemma A3-5:** In equilibrium the value of a high-cost $p^a$ is zero for any $\omega$: $V^a(\omega, 0) = 0$.

**Proof.** Assume that at state $\Omega = (\omega, 0)$ the ruler uses a strategy of the form $(\ell, HSS)$. Under such a strategy the offer to $p^a$ is such that a low-cost type supports with probability 1 and a high-cost type supports with interior probability. In such an equilibrium posterior beliefs are $\omega(A) > \omega$ and $\omega(R) = 0$. For a $\ell$ type,

$$m - \bar{\theta} + \delta V^a_\ell((\omega(A), 0)) = \delta V^a_\ell((0, 0))$$

Recall that $V^a_\ell((0, 0)) = 0$ so that the high-semi-separating offer must be $m = \bar{\theta} - \delta V^a_\ell((\omega(A), 0))$. But since a high-cost type cannot accept an offer below his cost of providing political support, and the equilibrium value of a politician cannot be negative, a HSS equilibrium can only exist if $V^a_\ell((\omega(A), 0)) = 0$, which implies $m = \bar{\theta}$. This is, the value at posterior beliefs $\Omega = (\omega(A), 0)$ must be zero for a high-cost $p^a$, and implies the equilibrium value for a high-cost politician at $\Omega$ must be $V^a_\ell(\Omega) = \bar{\theta} - \bar{\theta} + 0 = 0$.

Now assume at state $\Omega = (\omega, 0)$ the ruler uses a strategy of the form $(a, PA)$. Under such a strategy the ruler first makes a pooling offer to $p^a$, which both types of politician accept. If $p^a$ rejects the ruler makes an offer to $p^b$ which he accepts. Under a Pooling equilibrium $\omega(S) = \omega$. The high-cost type’s IC is

$$m - \bar{\theta} + \delta V^a_\ell(\Omega) \geq \delta V^a_\ell(\Omega)$$

where I assume that out-of-equilibrium beliefs after a rejection are $\omega(R) = \omega$ (the specific assumption about out-of-equilibrium beliefs is inconsequential). The IC above implies that the optimal offer in a pooling equilibrium when $p^a$ is made the first offer is $m = \bar{\theta}$. Of course, such an offer makes the low-cost type’s IC slack, and $V^a_\ell(\Omega) = 0$.

Finally, any other type of offer under $\Omega$ is either $(\ell, LSS)$ or $(\ell, PR)$. In either case the high-cost politician gets no transfer in the current period. Together these imply that $V^a_\ell(\Omega) = 0 \forall \Omega$ such that $\Omega = (\omega, 0)$.

**Lemma A3-6:** There are no HSS equilibrium offers.

**Proof.** First I show that if at $\Omega = (\omega, 0)$ the ruler makes a $LSS$ offer, then $V^a_\ell(\Omega) \leq \delta(\bar{\theta} - \bar{\theta})$. To show this observe that under an $LSS$ offer to $p^a$, a high-cost type must be rejecting, and a low-cost type mixes between acceptance and rejection. Posterior beliefs are $\omega(A) = 1$ and $\omega(R) < \omega$. The IC for the high-cost type is

$$\delta V^a_\ell((\omega(R), 0)) \geq m - \bar{\theta} + \delta V^a_\ell((1, 0)) = m(\Omega) - \bar{\theta}$$
The offer must also satisfy the low-cost type’s indifference condition:

\[ V^a_\theta((1,0)) = m(\Omega) - \theta + \delta V^a_\theta((1,0)) = \delta V^a_\theta((\omega(R),0)) \]

We have that \( V^a_\theta((1,0)) = 0 \) (see Appendix 2), so the offer must be \( m(\Omega) = \theta + \delta V^a_\theta((\omega(R),0)) \). At \((\omega(R),0)\) the ruler can make a \( PA \) offer, a \( PR \) offer, a \( LSS \) offer or a \( HSS \) offer.

Suppose the equilibrium offer at \((\omega(R),0)\) is \( PR \) (the ruler makes an offer to \( p^b \) and then no offer to \( p^a \)). Then the ruler makes no offer to \( p^a \) in any future period and there is no learning. In such case \( V^a_\theta((\omega(R),0)) = 0 \). This implies today’s offer must be \( m(Q) = \delta V^a_\theta((\omega(R),0)) \).

Now suppose the equilibrium offer at \((\omega(R),0)\) is \( PA \). Then the ruler makes a PA offer, a PR offer, a LSS offer or a HSS offer.

Suppose the equilibrium offer at \((\omega(R),0)\) is \( PR \). Then the ruler makes no offer to \( pa \) in any future period and there is no learning. In such case \( V^a_\theta((\omega(R),0)) = 0 \).

This implies today’s offer must be \( m(Q) = \delta V^a_\theta((\omega(R),0)) \).

Now consider a state \( \Omega \), and suppose that under \( \Omega \) \( HSS \) is an equilibrium. From Lemma A3-6, \( m = \bar{\theta}, \) and from Proposition 4 \( V^a_\theta((0,0)) = \frac{1}{2} \delta \bar{\theta} \delta \bar{\theta} \), so the low-cost type’s IC is

\[ \bar{\theta} - \theta + \delta V^a_\theta((\omega(A),0)) \geq \delta \frac{1}{2} \theta - \frac{1}{2} \theta \]

We conclude that if at \( \Omega \) the ruler makes an offer which induces \( LSS \), along an equilibrium, after a rejection -when posterior beliefs are \((\omega(R),0)\) - the ruler can only be making a \( PR \) or another \( LSS \) offer, and the low-cost type’s equilibrium value is bounded above by \( \delta (\bar{\theta} - \theta) \).
For an HSS offer to be an equilibrium in state Ω it is necessary that the above inequality holds. Since $V^1_e((ω(A),0)) = 0$ if at $(ω(A),0)$ the ruler makes a PR offer, in equilibrium a HSS offer cannot be followed by a PR offer. Since $V^1_e((ω(A),0)) < δ(\bar{θ} - \bar{θ})$ if at $(ω(A),0)$ the ruler makes a LSS offer, in equilibrium a HSS offer cannot be followed either by a LSS offer for large discount factors. Thus, a HSS offer can only be followed by a HSS offer or by a PA offer after support is observed. In either case, $m(ω(S),0) = \bar{θ}$.

Now, under a PA equilibrium the ruler offers $\bar{θ}$ to $p^θ$ every period, and beliefs remain unchanged. On the other hand, under a HSS equilibrium the ruler offers $\bar{θ}$ to $p^θ$, which, if rejected, drives the posterior to $(ω(R),0) = (0,0)$. In every future period the ruler must offer $\bar{θ}$ to one politician. If the offer is accepted, because an HSS offer can only be followed by a PA or another HSS offer, the ruler will also have to make an offer of $\bar{θ}$ to $p^θ$. Moreover, if at any period where the ruler is making a HSS offer to $p^θ$ first the outcome is a rejection, the ruler will have to make an offer of $\bar{θ} + \frac{1}{2}δ\phi$ to $p^θ$ to remain in power. Thus, a HSS offer cannot be an equilibrium; the ruler would prefer to make PS offer.

Lemma A3-6 above simply shows that high-cost semi-separation cannot be part of the equilibrium dynamics because for it to be feasible the ruler must make an offer as high as the offer which is required to induce pooling. At the same time, either “bad” information is revealed after a rejection -that $p^θ$ is high cost- or the ruler must keep making the same offer after support, while a rejection increases the value for $p^θ$ to reject, which is undesirable for the ruler.

Taken together the above lemmas imply that for $Ω = (ω,0)$ equilibrium strategies must be either $(a, PA)$ $(b, PR)$ or $(i, LSS)$. Moreover, that if at $Ω$ the ruler makes a LSS offer to $p^θ$, the offer cannot be $(a, PA)$ after a rejection (under beliefs $(ω(R),0)$). The ruler either cannot learn or can only induce partial separation. Recall from above that along an equilibrium in which the ruler makes an offer at $Ω$ and at $(ω(R),0)$ which induces LSS, the following must be satisfied:

$$m(ω(R),0) = \frac{1}{δ} m(Ω) - \frac{1}{δ} \bar{θ}$$

This equation has a unique fixed point at $\bar{θ}$, which is the lowest acceptable transfer the ruler can make which a low-cost politician can accept. Thus, offering $m = \bar{θ}$ satisfies the indifference condition for the low-cost type, and yields the highest possible payoff for the ruler when the equilibrium offer is LSS.

Notice that $(a, PA)$ and $(a, LSS)$ are “commitment” strategies in that $p^θ$ will never be pivotal because the ruler will make a second offer to $p^θ$ which accepts for sure whenever $p^θ$ rejects. On the other hand, under a strategy where $p^θ$ is made the first offer, if after a rejection by $p^θ$ the ruler has incentives not to make a pooling offer, $p^θ$’s outside option will be higher than his cost, and the ruler would have to make him a higher offer on the equilibrium path. Because the ruler’s value is sure to be $\frac{ρ-θ}{1-δ}$ under strategy $(a, PA)$, analogous to the derivation of the cutoff for beliefs $Ω = (1, ω)$, there exists a $ω$ such that for higher values the ruler would prefer to make a separating offer (a separating offer is feasible off the equilibrium path). The cuttof is given by the value of $ω$ for which the ruler, after a rejection by $p^θ$, prefers to make a separating rather than a pooling offer to $p^θ$.

The interim value for a ruler who makes a pooling offer $m^2 = \bar{θ} + \frac{1}{2}δ\phi$, just enough for the high-cost type to accept, is $v'(Ω) = μ - \bar{θ} - \frac{1}{2}δ\phi + δW'(Ω)$, where notice that beliefs remain unchanged given that both types of $p^θ$ accept. A separating offer must be just enough for the low-cost type to support: $m^2 = \bar{θ} + \frac{1}{2}δ\phi$ (this offer is rejected by the high-cost type). By making a separating offer the interim value for the ruler is $v'(Ω) = μ + ω[δW'(1,0) - \bar{θ} - \frac{1}{2}δ\phi]$. Pooling off-the-equilibrium path is credible iff

$$ω < \frac{δW'(ω,0) - \bar{θ} - \frac{1}{2}δ\phi}{δW'(ω,0) - \bar{θ} - \frac{1}{2}δ\phi} = \hat{ω}(δ)$$

151
Given that in any equilibrium under a state \( \Omega = (\omega, 0) \) the ruler’s value is bounded between \( \frac{\mu - \bar{\theta}}{1-\delta} \) and \( \frac{\mu - \bar{\theta}}{1-\delta} \), there exists a cutoff \( \hat{\omega} \in \left[ \frac{(\mu - \bar{\theta}) - (1-\delta)\delta}{1-\delta} \frac{(\mu - \bar{\theta}) - (1-\delta)\delta}{1-\delta} \right] \). Notice the upper bound of this set is \( \bar{\omega} \) as defined in Lemma A3-2. Since a strategy \((a, PA)\) gives the ruler a value he cannot attain under strategy \((b, PR)\) for \( \omega > \hat{\omega} \), there exists some \( \omega' > \hat{\omega} \) such that the ruler will make a first offer to \( p^a \) for any \( \omega > \omega' \).

**Lemma A3-7:** There exist \( \hat{\omega} \leq \hat{\omega} \) such that for \( \omega < \hat{\omega} \) the ruler prefers a strategy \((b, LSS)\) (under commitment) to a strategy \((a, LSS)\).

**Proof.** The ruler’s equilibrium value of following a strategy \((b, LSS)\) is

\[
W^*_b(Q) = \mu - \bar{\theta} + \omega \sigma \left[ \delta \frac{\mu - \bar{\theta}}{1-\delta} - \bar{\theta} \right] + (1 - \omega \sigma) \delta W^* (\omega(\Omega), 0)
\]

which follows because the ruler first makes an offer \( \bar{\theta} \) to \( p^b \) which is accepted, and then makes an offer \( m^2 = \bar{\theta} \) to \( p^b \), which is accepted with probability \( \omega \sigma \) and rejected with probability \( 1 - \omega \sigma \). The ruler’s value of following a strategy \((a, LSS)\) is

\[
W^*_a(Q) = \mu - \bar{\theta} + \omega \sigma \left[ \delta \frac{\mu - \bar{\theta}}{1-\delta} - \bar{\theta} \right] + (1 - \omega \sigma) \delta W^* (\omega(\Omega), 0) - \bar{\theta} - \frac{1}{2} \delta \varphi
\]

which follows because the ruler first makes a \( LSS \) offer \( m^1 = \bar{\theta} \) to \( p^b \) which is accepted with probability \( \omega \sigma \) and rejected with probability \( 1 - \omega \sigma \). After a rejection, the ruler must make a second offer \( m^2 = \bar{\theta} + \frac{1}{2} \delta \varphi \) for \( p^b \) to support. Take the limit as \( \omega \to 0 \). Given that \( \sigma \) is bounded between 0 and 1, \( \omega \sigma \to 0 \) too. It follows that

\[
\lim_{\omega \to 0} W^*_a(Q) < \lim_{\omega \to 0} W^*_b(Q)
\]

Thus, \( \omega > 0 \) exists.

Lemma A3-7 highlights the trade-off between making a first offer to \( p^a \) or to \( p^b \). The former implies that on the equilibrium path the ruler might have to make a large offer to a pivotal politician, but is a minimum winning coalition. The latter makes no politician pivotal, but makes a supermajority coalition possible on the equilibrium path.

\[\square\]

**Lemma A3-8:** For any \( \omega < \hat{\omega} \) and any mixing probability by a low-cost politician, the ruler prefers \((b, LSS)\) to \((b, PR)\).

**Proof.** For any \( \Omega \in \{ (\omega, 0) : \omega \in (0, 1) \} \), the ruler’s equilibrium value is bounded below by \( \frac{\mu - \bar{\theta}}{1-\delta} \), so that

\[
W^*_b(Q) = \mu - \bar{\theta} + \omega \sigma \left[ \delta \frac{\mu - \bar{\theta}}{1-\delta} - \bar{\theta} \right] + (1 - \omega \sigma) \delta \frac{\mu - \bar{\theta}}{1-\delta} = \mu - \bar{\theta} + \omega \sigma \left[ \delta \frac{\bar{\theta} - \bar{\theta}}{1-\delta} - \bar{\theta} \right]
\]

This is larger than \( W^*_b(\Omega) = \frac{\mu - \bar{\theta}}{1-\delta} \) for any \( \delta \) large enough \( \delta \geq \bar{\theta}/\bar{\theta} \).

\[\square\]

**Equilibrium cutoff \( \hat{\omega} \) for \( \sigma^a = 1/2 \).**

Assume \( \Omega = (0, \omega^b) \), so that \( p^a \)’s type is known to be high-cost. Recall equilibrium behavior for \( \omega^b < \hat{\omega} \) prescribes the ruler to make a first offer \( m^1 = \bar{\theta} \) to \( p^a \), the politician whose type is common knowledge, which \( p^a \) accepts. The ruler then makes a second offer \( m^2 = \bar{\theta} \) to \( p^b \), which a high-cost \( p^b \) rejects and a low-cost \( p^b \) accepts with probability 1/2. Posterior beliefs after \( k \) rejections are thus
\[ \omega_k^b = \frac{\omega_k^b}{2^k - (2^k - 1)\omega_k^b} \]

Thus, the ruler’s ex-ante value can be expressed recursively as

\[
W^R(0, \omega_k^b) = \mu - \theta - \frac{1}{2}\delta \omega_k^b + \frac{1}{2}\delta \omega_k^b \delta W^R(0,1) + [1 - \frac{1}{2}\delta \omega_k^b] \delta W^R(0, \frac{\omega_k^b}{2 - \omega_k^b})
\]

Iteratively,

\[
W^R(0, \omega_k^b) = \mu - \theta + \frac{1}{2}\omega_k^b \left[ \delta \frac{\mu - \theta}{1 - \delta} - \theta \right] + [1 - \frac{1}{2}\omega_k^b] \delta \left[ \mu - \theta + \frac{1}{2}\omega_k^b \left[ \delta \frac{\mu - \theta}{1 - \delta} - \theta \right] \right]
\]

\[+ [1 - \frac{1}{2}\omega_k^b] \delta \left[ \mu - \theta + \frac{1}{2}\omega_k^b \left[ \delta \frac{\mu - \theta}{1 - \delta} - \theta \right] \right] + [1 - \frac{1}{2}\omega_k^b] \delta \left[ \mu - \theta + \frac{1}{2}\omega_k^b \left[ \delta \frac{\mu - \theta}{1 - \delta} - \theta \right] \right] + \ldots \]

which can be reorganized as

\[
W^R(0, \omega_k^b) = \left[ \mu - \theta \right] (1 + \delta T_1 + \delta^2 T_2 + \delta^3 T_3 + \ldots) + \left[ \delta \frac{\mu - \theta}{1 - \delta} - \theta \right] \left( 1 - T_1 + \delta T_1 \frac{1}{2}\omega_k^b + \delta^2 T_2 \frac{1}{2}\omega_k^b + \delta^3 T_3 \frac{1}{2}\omega_k^b + \ldots \right)
\]

where \( T_k = \frac{1}{2^k} \omega_k^b \). Solving the equation above,

\[
W^R(0, \omega_k^b) = \left[ \mu - \theta \right] \frac{2 - \delta(1 + \omega_k^b)}{(1 - \delta)(2 - \delta)} + \left[ \delta \frac{\mu - \theta}{1 - \delta} - \theta \right] \frac{\omega_k^b}{2 - \delta}
\]

\[
W^R(0, \omega_k^b) = \frac{\mu - \theta}{1 - \delta} + \frac{\omega_k^b}{2 - \delta} \left[ \delta \frac{\mu - \theta}{1 - \delta} - \theta \right]
\]

Notice that for \( \delta \geq \theta/\bar{\theta} \), the ruler’s equilibrium value is increasing in \( \omega_k^b \). Of course, when the ruler faces no commitment problem following a rejection by \( p^1 \), his equilibrium value should be increasing in the degree of optimism about \( p^2 \)’s type. The term multiplying \( \omega_k^b \) in fact represents the expected present value of informational rents that the ruler will save from transferring to \( p^a \) in case \( p^b \) reveals his type is \( \theta \). In fact, equilibrium values for both politicians along the equilibrium are \( V_a^b(0, \omega_k^b) = 0 \), \( V_{\theta}^b(0, \omega_k^b) = 0 \), \( V_{\theta}^a(0, \omega_k^b) = 0 \), \( V_{\theta}^a(0, \omega_k^b) = \frac{\theta - \gamma}{1 - \delta} - \omega_k^b \left[ \frac{\delta(\theta - \gamma)}{1 - \delta} \right] \).

\( \gamma \) is given by beliefs such that after a rejection by \( p^a \), the ruler’s interim value of inducing \( S \) is higher than the value of following the equilibrium prescribing \( LSS \) to \( p^b \). Hence, \( \gamma \) satisfies:

\[
\mu + \omega_k^b \left[ \frac{\delta(\theta - \gamma)}{1 - \delta} \right] = \mu - \theta - \frac{1}{2}\delta \phi + \delta \frac{\mu - \theta}{1 - \delta} + \frac{\omega_k^b}{2 - \delta} \left[ \frac{\delta(\theta - \gamma)}{1 - \delta} \right]
\]

\[
\Leftrightarrow
\]
\[ \dot{\omega} = \frac{\delta(\mu - \bar{\theta}) - (1 - \delta)\bar{\theta} - \delta(1 - \delta)\frac{1}{2}\phi}{\delta(\mu - \bar{\theta}) - (1 - \delta)\bar{\theta} - \delta(1 - \delta)\frac{1}{2}\phi - \delta\frac{\Delta\bar{\theta} - \bar{\phi}}{2 - \delta}} \]
Appendix 4: Interior Beliefs

Proposition 8: Entrenchment Dynamics at $\Omega = (\omega^a, \omega^b)$.

I start by proving the following lemma:

Lemma A4-1: For any interim beliefs $\Omega$, where $D^1 = R$, an $S$ offer is always feasible.

Proof. Without loss of generality assume $p^a$ has rejected, and interim beliefs are $(\omega^a(R), \omega^b)$. The low-cost type incentive constraint for separation is:

$$m^2 - \theta + \delta V^b_{\theta}(\omega^a(R), 1) \geq \frac{1}{2} \phi$$

For the high-cost type the incentive constraint is:

$$m^2 - \bar{\theta} + \delta V^b_{\bar{\theta}}(\omega^a(R), 1) < \frac{1}{2} \phi$$

These ICs follow from the fact that in a separating equilibrium, acceptance drives the posterior to 1, while rejection triggers a power transition. Reducing $m^2$ until the low type’s IC binds increases the ruler’s value, and relaxes the high type’s IC. Thus, $m^2 = \max \{\theta + \delta \frac{1}{2} \phi - \delta V^b_{\theta}(\omega^a(R), 1, \theta), 0\}$. At $\Omega = (\omega^a(R), 1)$, $V^b_{\theta}(\omega^a(R), 1) = 0$ if $\omega^a(R) < \bar{\omega}$, in which case $m^2 = \theta + \frac{1}{2} \phi$, or $V^b_{\theta}(\omega^a(R), 1) = (1 - \omega^a(R))\delta \frac{1}{2} \phi$ if $\omega^a(R) \geq \bar{\omega}$, in which case $m^2 = \theta + \frac{1}{2} \phi [1 - \delta (1 - \omega^a(R))]$.

If $\omega^a(R) < \bar{\omega}$, $V^b_{\theta}(\omega^a(R), 1) = 0$ given that the ruler will make positive offers of exactly $\theta$ to $p^b$ in the future. Hence, a separating equilibrium is feasible iff

$$\theta + \frac{1}{2} \phi - \bar{\theta} < \frac{1}{2} \phi \iff \bar{\theta} - \theta < 0$$

which is of course satisfied.

If $\omega^a(R) \geq \bar{\omega}$, at $\Omega = (\omega^a(R), 1)$ the ruler makes a first separating offer to $p^a$, which is accepted if $p^a$ is low cost. If $p^a$ rejects, the ruler makes an offer $m^2 = \theta + \delta \frac{1}{2} \phi$ to $p^b$, which $p^b$ rejects, and a power transition ensues. Thus, $V^b_{\theta}(\omega^a(R), 1) = \omega^a(R) \cdot 0 + (1 - \omega^a(R))\delta \frac{1}{2} \phi = (1 - \omega^a(R))\delta \frac{1}{2} \phi$. Hence, a separating equilibrium is feasible iff

$$\theta + \delta \frac{1}{2} \phi [1 - \delta (1 - \omega^a(R))] - \bar{\theta} + \delta^2 (1 - \omega^a(R)) \frac{1}{2} \phi < \frac{1}{2} \phi \iff \bar{\theta} - \theta < 0$$

which is also satisfied. $\square$

The ruler can always separate after a rejection, because following a rejection the outside option of both types of politician is the same, given that the value of triggering a political transition does not depend on the
politician's type. Lemmas A4-2 and A4-3 now characterize equilibrium behavior at an interim stage in which revelation has taken place.

**Lemma A4-2:** At Interim beliefs \((1, \omega)\) after \(D^1 = A\), the ruler's offer to \(b\) is \(PR\).

**Proof.** First, assume that \(\omega < \bar{\omega}\). Trivially, making any kind of offer to \(p^2\) is suboptimal for the ruler, who will already remain in power achieving his maximum value.

Now assume \(\omega \geq \bar{\omega}\) notice that if interim beliefs are \((1, \omega)\), and the ruler made a first offer to \(p^2\), it must be that \(a\) accepted. The ruler will already remain in power until next period. The ruler could make an informative offer to \(p^b\), or wait until next period and follow the equilibrium behavior at \((1, \omega)\). In the latter case, the ruler's interim value in the current period is

\[
v^r(1, \omega) = \mu - m^1 + \delta W^r(1, \omega) = \mu - m^1 + \delta \left[ \frac{\mu - \theta}{1 - \delta} - (1 - \omega)\frac{1}{2}\delta \right]
\]

First notice that for interim beliefs \((1, \omega)\), the ruler can actually make a fully separating offer of \(m^2 = \theta\). This follows because after either acceptance or rejection, next period the equilibrium value for both types of politician is zero. Of course, full revelation is strictly preferred to LSS or HSS by the ruler. The ruler's interim value of separation in the current period is

\[
v^r(1, \omega)|_s = \mu - m^1 + \omega \left[ -\theta + \delta \frac{\mu - \theta}{1 - \delta} \right] + (1 - \omega) \left[ \delta \frac{\mu - \theta}{1 - \delta} \right]
\]

This implies that separating at the interim stage is optimal iff \(v^r(1, \omega)|_s > v^r(1, \omega)\) \iff

\[
\omega < \frac{\frac{\delta^2}{2} \phi}{\bar{\theta} + \frac{\delta}{2} \phi}
\]

Now, for \(\delta\) large enough, \(\frac{\delta^2}{2} \phi < \bar{\omega}\) so the ruler prefers to \(PR\) for any \(\omega \in [\bar{\omega}, 1)\) (for \(\delta \geq \bar{\theta}/\bar{\theta}\) this inequality is slack). Of course, an analogous argument applies for interim beliefs \((\omega, 1)\).

From Lemma A4-2 it follows that starting from beliefs \((\omega^a, \omega^b)\), following a fully revealing acceptance decision by \(a\) that drives interim posterior beliefs to \((1, \omega^b)\), the ruler will not make a second offer to \(b\).

**Lemma A4-3:** At Interim beliefs \((0, \omega)\), such that \(\omega \geq \bar{\omega}\), the ruler's offer to \(b\) is \(S\). If \(\omega < \bar{\omega}\), the ruler's offer to \(b\) is \(PS\).

**Proof.** Interim beliefs at \((0, \omega)\) must have followed a revealing rejection by \(p^a\). Thus, \(p^b\) is pivotal. From Lemma A4-1, we know separation is feasible. Now observe that \(\bar{\omega}\) is precisely defined as the maximum value of beliefs about \(p^b\) for which the ruler prefers not to separate following a rejection. It thus follows that that for \(\omega \geq \bar{\omega}\), the ruler will make a separating second offer, while for \(\omega < \bar{\omega}\) the ruler will make a pooling second offer.

Of course, a symmetric arguments applies for beliefs \((\omega, 0)\) following a rejection by \(p^b\).

156
Lemmas A4-4, and A4-5 now show that for interior beliefs, the ruler cannot make a fully separating first offer if beliefs about the second politician to be offered are $w < \tilde{w}$, and that after acceptance by $p^1$ the ruler cannot make a fully separating second offer if beliefs about the second politician to be offered are $\omega < \tilde{\omega}$. This will restrict considerably the type of equilibria that can be sustained.

**Lemma A4-4:** For any interior beliefs $\Omega = (\omega_a, \omega_b)$, a separating first offer to $a$ is infeasible if $\omega^b \leq \tilde{\omega}$, and feasible for $\omega^b > \tilde{\omega}$.

**Proof.** From Lemmas A4-2 and A4-3, it follows that there are two relevant regions of the state space concerning a separating offer to $p^1$. Without loss of generality assume that $p^1 = a$. After a separating offer to $p^a$, posterior beliefs become either $(1, \omega^b)$ or $(0, \omega^b)$, following an acceptance or a support decision, respectively. In region $\Omega_1 = \{(\omega^a, \omega^b) : \omega^b < \tilde{\omega}\}$, the ruler will make a pooling offer to $p^b$ following a rejection, while in region $\Omega_2 = \{(\omega^a, \omega^b) : \omega^b \geq \tilde{\omega}\}$ the ruler will make a separating offer to $p^b$ following a rejection (see Lemma A4-3). In either region, the ruler will make no offer to $p^b$ following an acceptance decision.

Assume beliefs are $\Omega \in \Omega_1$. Incentive constraints for a separating equilibrium in this region are, for a low-cost type,

\[ m^1 - \theta + \delta V^a_{\tilde{\theta}}(1, \omega^b) \geq \delta V^a_{\tilde{\theta}}(0, \omega^b) \]

and for a high-cost type,

\[ m^1 - \bar{\theta} + \delta V^a_{\bar{\theta}}(1, \omega^b) < \delta V^a_{\bar{\theta}}(0, \omega^b) \]

The ICs follow from the fact that after a rejection, the ruler will pool $p^b$, and after acceptance the ruler will make no offer to $p^b$. Moreover, in this region we have that $V^a_{\tilde{\theta}}(1, \omega^b) = V^a_{\tilde{\theta}}(1, \omega^b) = V^a_{\bar{\theta}}(0, \omega^b) = 0$, while $V^a_{\bar{\theta}}(0, \omega^b) = \tilde{\theta} - \theta + \delta (1 - \sigma^b \omega^b)V^a_{\bar{\theta}}(0, \omega^b(R))$. Reducing $m^1$ until the low type’s IC binds increases the ruler’s value, and relaxes the high type’s IC. Thus, $m^1 = \theta + \delta (\bar{\theta} - \theta) + \delta (1 - \sigma^b \omega^b)V^a_{\bar{\theta}}(0, \omega^b(R))$. Separation is feasible iff under this offer the low-cost politician rejects:

\[ \delta^2 (1 - \sigma^b \omega^b) V^a_{\bar{\theta}}(0, \omega^b(R)) < (\bar{\theta} - \theta)(1 - \delta) \]

This condition cannot be satisfied for large enough $\delta$, for any mixing probability of a low-cost $p^b$. For the case in which $\sigma^b = 1/2$ (see Appendix 3 above) for example, the condition is

\[ (\bar{\theta} - \theta) \left\{ \delta \left[ 1 + \left( \frac{1}{1 - \delta} - \frac{\omega^b}{2 - \omega^b} \left[ \frac{\delta}{2 - \delta} \right] \delta (1 - \sigma^b \omega^b) \right] - 1 \right\} < 0 \]

which would only be satisfied for very small $\delta$.

Now assume beliefs are $\Omega \in \Omega_2$. Incentive constraints for a separating equilibrium in this region are, for a low-cost type,

\[ m^1 - \theta + \delta V^a_{\tilde{\theta}}(1, \omega^b) \geq \omega^b \delta V^a_{\tilde{\theta}}(0, 1) + (1 - \omega^b) \delta \frac{1}{2} \phi \]

and for a high-cost type,

\[ m^1 - \bar{\theta} + \delta V^a_{\bar{\theta}}(1, \omega^b) < \omega^b \delta V^a_{\bar{\theta}}(0, 1) + (1 - \omega^b) \delta \frac{1}{2} \phi \]
The ICs follow from the fact that in this region, following a rejection by \( p_1 \) the ruler makes a separating offer to \( p_2 \) (see Lemma A4-3), and in case \( p_2 \) rejects, there is a power transition. We have that \( V^{a}_{\hat{\theta}}(0,1) = V^{a}_{\hat{\theta}}(0,1) = 0 \). If \( \omega^b < \hat{\omega} \), \( V^{a}_{\hat{\theta}}(1,\omega^b) = V^{a}_{\hat{\theta}}(1,\omega^b) = 0 \), so that \( m^1 = \hat{\theta} + (1 - \omega^b)\delta_{1/2}\phi \) satisfies the low-cost type politician. Separation is feasible iff

\[
\hat{\theta} + (1 - \omega^b)\delta_{1/2}\phi - \hat{\theta} < (1 - \omega^b)\delta_{1/2}\phi \Leftrightarrow \\
\hat{\theta} - \hat{\theta} < 0
\]

which is trivially satisfied. If \( \omega^b \geq \hat{\omega} \), on the other hand, \( V^{a}_{\hat{\theta}}(1,\omega^b) = (1 - \omega^b)\delta_{1/2}\phi \) given that next period at \((1,\omega^b)\), the ruler will separate \( b \) and in case of a rejection will make a high enough offer \( m = \hat{\theta} + \delta_{1/2}\phi \) for \( a \) to accept, while \( V^{a}_{\hat{\theta}}(1,\omega^b) = (1 - \omega^b)\delta_{1/2}\phi \), since a high-cost \( a \) will reject the offer and a power transition takes place. For separation in the current period, the ruler will offer \( m^1 = \hat{\theta} + (1 - \omega^b)\delta_{1/2}\phi - (1 - \omega^b)\delta_{1/2}\phi \). Separation is feasible iff

\[
\hat{\theta} + (1 - \omega^b)\delta_{1/2}\phi - (1 - \omega^b)\delta_{1/2}\phi - \hat{\theta} + (1 - \omega^b)\delta_{1/2}\phi < (1 - \omega^b)\delta_{1/2}\phi \Leftrightarrow \\
\hat{\theta} - \hat{\theta} < 0
\]

which is again trivially satisfied. \( \square \)

Lemma A4-4 shows that Separation is infeasible when beliefs about the second politician are too pessimistic, because the upwards IC cannot be satisfied. While after low-type revelation the ruler will fully entrench and encroach, a low-cost politician has strong incentives to be believed high cost, and collect informational rents in the future as long as the ruler keeps testing \( p^t \). Hence, inducing separation is too costly, and requires a current transfer that a high cost politician would be ready to accept. Thus, when in the future a ruler is willing to reduce his encroachment to be able to entrench by giving rents, this limits his present ability to learn because it gives incentives for low-cost politicians to try to capture those future rents. In contrast, when the ruler is willing to incur some risk by making very informative offers, currently he can more easiliy learn because separation is made easier when the outside option of both types of politician is the same, in this case, a power transition. This only happens when beliefs about \( p^2 \) are high enough.

**Lemma A4-5:** There exists \( \hat{\omega} > \tilde{\omega} \) such that for any interim beliefs \( \Omega = (\omega_a, \omega_b) \) with \( \omega_a < \tilde{\omega} \), and where \( D^p = A \), an \( S \) offer to \( b \) cannot be part of an equilibrium. Moreover, for \( \omega_a \geq \hat{\omega} \), an \( S \) offer to \( b \) is feasible.

**Proof.** Separation at the interim stage requires, for a low-cost politician \( b \),

\[
m_2 - \hat{\theta} + \delta V^b_{\hat{\theta}}(\omega_a,1) \geq V^b_{\hat{\theta}}(\omega_a,0)
\]

and for a high-cost, politician \( b \).
Notice that $V_0^b(w_a, 1) = V_0^b(w_a, 1)$, which is either 0 if $w_a < \bar{w}$, or $(1 - w_a)\delta \phi$ if $w_a \geq \bar{w}$. Moreover, $V_0^b(w_a, 0) = 0$. The ruler will offer $m_2 = \max\{\bar{\theta} + \delta V_0^b(w_a, 0) - \delta V_0^b(w_a, 1), \theta\}$, and separation will be feasible iff

$$\delta V_0^b(w_a, 0) < \bar{\theta} - \theta$$

Recall $V_0^b(w_a, 0)$ is decreasing in $w_a$ on the interval $[0, \bar{w}]$. Thus, if the constraint above is not satisfied for $w_a = \bar{w}$ it will not be satisfied anywhere in the interval. For $\sigma^a = 1/2$, $V_0^b(w_a, 0) = \frac{\bar{\theta} - \delta}{1 - \delta} - w_a \frac{\delta \phi}{1 - \delta}$, so that the IC above can only be satisfied for $w_a > \frac{2 - \delta}{\delta(1 - \delta)} > 1$. Thus, separation is not feasible for any $w_a < \bar{w}$.

Now observe for a given support strategy of $a$ along boundary $(w_a, 0)$, $V_0^b(w_a, 0)$ is continuous and decreasing in the interval $[\bar{w}, 1]$, and $\lim_{w_a \to 1} V_0^b(w_a, 0) = 0$. Thus, $\exists \bar{w}$ which satisfies $\delta V_0^b(\bar{w}, 0) = \bar{\theta} - \theta$, such that $\forall w_a \geq \bar{w}$, the incentive constraint for a high cost politician $b$ is satisfied, and separation is feasible.

Finally, notice that $V_0^b(w_a, 0) > V_0^b(\bar{w}, 0)$ for all $w_a > (\bar{w}, \bar{w})$, so that separation is also infeasible in this range. Thus, for $w_a < \bar{w}$ the ruler cannot separate $b$ at the interim stage, and for $w_a > \bar{w}$ the ruler can separate $b$ at the interim stage. □

For beliefs in the region $\Omega^1 = \{(w_a, w_b) : w_b < \bar{w}\}$, lemmas A4-4 and A4-5 imply that the ruler either cannot make a first fully separating offer to $a$, and cannot make a separating second offer to $a$ after $b$ has accepted. The following lemma shows that for any interior beliefs, a strategy that first makes a pooling offer to the politician under more pessimistic beliefs (to give incentives for the second politician to reveal his type), followed by a LSS offer to the politician under more optimistic beliefs, cannot be an equilibrium.

**Lemma A4-6:** For interior beliefs $(\omega^a, \omega^b)$ such that $\omega^a > \omega^b$, a strategy of the form $\gamma = (b, PA, LSS, \cdot)$ cannot be part of a MPBE.

*Proof.* Omitted. □

Thus, a strategy of the form $\gamma = (b, PA, LSS, \cdot)$ cannot be part of an equilibrium for any beliefs.

**Lemma A4-7:** Assume $\Omega = (\omega^a, \omega^b)$ such that $\omega^a \geq \omega^b$. For given $\omega^a$, strategy $(a, HSS, LSS, PA)$ is an equilibrium for any $\omega^b \leq \bar{w}$.

*Proof.* Omitted. □

**Lemma A4-8:** Assume $\Omega = (\omega^a, \omega^b)$ such that $\omega^a \geq \omega^b$. For given $\omega^a$, strategy $(a, PA, PR, PA)$ is an equilibrium for $\bar{w}(\omega^a) \geq \omega^b > \bar{w}$.

*Proof.* Omitted. □

Now I show that when the optimal offer strategy for the ruler is to make a first separating offer and a second separating offer following a rejection, then he must make the first offer to the politician under more pessimistic beliefs.
Lemma A4-9: For beliefs $\Omega$ such that $\gamma = (\ell, S_\ell, PR_\ell, S_\ell)$, $\ell$ is such that $\omega^\ell < \omega^{\ell'}$.

Proof. The equilibrium value of strategy $\gamma$ for the ruler is given by

\[
W^R(\Omega) = \mu + \omega_1 \left[ -m_1 + \delta W^R(1, \omega_2) \right] + (1 - \omega_1) \left[ \omega_2 (-m_2 + \delta W^R(0, 1)) + (1 - \omega_2) \cdot 0 \right] \\
= \mu + \omega_1 \left[ \delta \left( \frac{\mu - \theta}{1 - \delta} - (1 - \omega_2)\delta \frac{1}{2} \phi \right) - m_1 \right] + (1 - \omega_1)\omega_2 \left( \delta \frac{\mu - \theta}{1 - \delta} - m_2 \right)
\]

which follows directly from the ruler’s equilibrium values at the boundaries (see Propositions 6 and 5). Following a rejection, a separating offer for $p^2$ must be $m_2 = \theta + \delta \frac{1}{2} \phi$. On the other hand, the Incentive Constraints for separation of $p^1$ are, for a low-cost politician,

\[
m_1 - \theta + \delta V^2_2 (1, \omega_2) \geq \omega_2 \delta V^1_2 (0, 1) + (1 - \omega_2)\delta \frac{1}{2} \phi
\]

and for a high-cost politician,

\[
m_1 - \bar{\theta} + \delta V^2_2 (1, \omega_2) < \omega_2 \delta V^1_2 (0, 1) + (1 - \omega_2)\delta \frac{1}{2} \phi
\]

This implies $m_1 = \theta + (1 - \omega_2) \delta \frac{1}{2} \phi$, and the high-cost’s IC is slack. Replacing equilibrium offers in the ruler’s value,

\[
W^R(\Omega) = \mu + \omega_1 \left[ \delta \left( \frac{\mu - \theta}{1 - \delta} - (1 - \omega_2)\delta \frac{1}{2} \phi \right) - \left( \theta + (1 - \omega_2)\delta \frac{1}{2} \phi \right) \right] + (1 - \omega_1)\omega_2 \left( \delta \frac{\mu - \theta}{1 - \delta} - (\theta + \delta \frac{1}{2} \phi) \right)
\]

Notice the term $[\omega_1 + \omega_2 - \omega_1 \omega_2]$ takes the same value irrespective of $\omega_1 > \omega_2$ or $\omega_1 < \omega_2$. On the other hand, the term $[(1 - \omega_2)\omega_1 (1 + \delta) + \omega_2 (1 - \omega_1)]$ is not. In fact,

\[
[(1 - \omega_2)\omega_1 (1 + \delta) + \omega_2 (1 - \omega_1)] < [(1 - \omega_1)\omega_2 (1 + \delta) + \omega_1 (1 - \omega_2)] \\
[\omega_1 (1 + \delta) - \omega_2 \omega_1 (1 + \delta) + \omega_2 - \omega_2 \omega_1] < [\omega_2 (1 + \delta) - \omega_1 \omega_2 (1 + \delta) + \omega_1 - \omega_1 \omega_2] \\
[\omega_1 (1 + \delta) + \omega_2] < [\omega_2 (1 + \delta) + \omega_1] \\
\omega_1 < \omega_2
\]

which implies that it is optimal for the ruler to make the first offer to the politician which he is more pessimistic about. \qed

Lemma A4-10: Assume $\Omega = (\omega^a, \omega^b)$ such that $\omega^a \geq \omega^b$. For given $\omega^a$, strategy $(b, S, PR, S)$ is an equilibrium for $\omega^b > \bar{\omega}(\omega^a)$.

Proof. Omitted. \qed
Lemma A4.11: Cutoff \( \hat{\omega} (\omega^a) \): The cutoff \( \hat{\omega} (\omega^a) \) is implicitly given by equation 14

Proof. Assume for now that \( \omega^a < \bar{\omega} \). The ruler must compare following a strategy \((a, PA, PR, PA)\) with a strategy \((b, S, PR, S)\). For the latter strategy, after a rejection by \( b \) the incentive constraint for a low-cost type \( a \) is

\[
m^2 - \bar{\theta} + \delta V^b_\theta (1,0) \geq \delta \frac{1}{2} \phi
\]

For a high-cost type to reject, it must be that

\[
m^2 - \bar{\theta} + \delta V^b_\theta (1,0) < \delta \frac{1}{2} \phi
\]

which implies an equilibrium second offer of \( m^2 = \bar{\theta} + \delta \frac{1}{2} \phi \), given that \( V^b_\theta (1,0) = V^b_\theta (1,0) = 0 \)

Now, a first separating offer to \( b \) must satisfy the following incentive constraints: For a low-cost type,

\[
m^1 - \bar{\theta} + \delta V^b_\theta (\omega^a,1) \geq \omega^a \delta V^b_\theta (1,0) + (1 - \omega^a) \delta \frac{1}{2} \phi
\]

and for a high-cost type,

\[
m^1 - \bar{\theta} + \delta V^b_\theta (\omega^a,1) < \omega^a \delta V^b_\theta (1,0) + (1 - \omega^a) \delta \frac{1}{2} \phi
\]

which implies an equilibrium first offer of \( m^1 = \bar{\theta} + (1 - \omega^a) \delta \frac{1}{2} \phi \), given that \( V^b_\theta (\omega^a,1) = V^b_\theta (1,0) = 0 \) (Observe that if \( \omega^a \geq \bar{\omega} \), then \( V^b_\theta (\omega^a,1) = (1 - \omega^a) \delta \frac{1}{2} \phi \), and it would be even easier to separate.)

The ruler’s ex-ante value is then

\[
W^r (\Omega) |_{(b,S,PR,S)} = \mu + \omega^b \left[ -\bar{\theta} - (1 - \omega^a) \delta \frac{1}{2} \phi + \delta \frac{\mu - \bar{\theta}}{1 - \delta} \right] + (1 - \omega^b) \omega^a \left[ -\bar{\theta} - (1 - \omega^a) \delta \frac{1}{2} \phi + \delta \frac{\mu - \bar{\theta}}{1 - \delta} \right]
\]

Thus, \((a, PA, PR, PA)\) is prefered by the ruler to \((b, S, PR, S)\) if

\[
\frac{\mu - \bar{\theta}}{1 - \delta} > \mu + \omega^b + (1 - \omega^b) \omega^a \left[ \delta \frac{\mu - \bar{\theta}}{1 - \delta} - \bar{\theta} \right] - [\omega^b + \omega^a - 2\omega^b \omega^a] \delta \frac{1}{2} \phi
\]

This binding inequality implicitly defines \( \hat{\omega} (\omega^a) \) (It can be shown through outright differentiation that the condition \( \delta [\mu - (1 - \delta) \phi] - \bar{\theta} \geq 0 \) is sufficient for \( \hat{\omega} (\omega^a) \) to be decreasing and concave. Notice this condition can be interpreted as saying that a power transition is sufficiently economically costly. I omit it to save space).
References


Jones, Benjamin and Benjamin Olken, “Hit or Miss: The Effect of Assassinations on Institutions and War,” July 2008.


Chapter 3: The Myth of the Frontier*

Camilo Garcia-Jimeno†

James A. Robinson‡

June, 2011

Abstract

One of the most salient explanations for the distinctive path of economic and political development of the United States is captured by the “Frontier (or Turner) thesis”. Turner argued that it was the presence of the open frontier which explained why the United States became democratic and, at least implicitly, prosperous. In this paper we provide a simple test of this idea. We begin with the contradictory observation that almost every Latin American country had a frontier in the 19th century as well. We show that while the data does not support the Frontier thesis, it is consistent with a more complex “conditional Frontier thesis”. In this view, the effect of the frontier is conditional on the way that the frontier was allocated and this in turn depends on political institutions at the time of frontier expansion. We show that for countries with the worst political institutions, there is a negative correlation between the historical extent of the frontier and contemporary income per-capita. For countries with better political institutions this correlation is positive. Though the effect of the frontier on democracy is positive irrespective of initial political institutions, it is larger the better were these institutions. In essence, Turner saw the frontier as having positive effects on development because he already lived in a country with good institutions.

---

*This paper constitutes the third chapter of my Ph.D. dissertation at MIT, and was originally written for the conference Understanding Long-Run Economic Growth: A Conference Honoring the Contribution of Kenneth Sokoloff. Our biggest debt is to Ken for all his encouragement, friendship and inspiration over the years. We miss him. We also thank Giovanni Zambotti at the Center for Geographic Analysis at Harvard for his enormous help with the maps.

†MIT, Department of Economics. cgarcia@mit.edu. 50 Memorial Drive, Of. E52-201, Cambridge, MA 02139.

‡Harvard University, Department of Government. IQSS, 1737 Cambridge St., N309, Cambridge MA02138; e-mail: jrobinson@gov.harvard.edu.
1 Introduction

One of the great economic puzzles of the modern world is why, amongst a group of colonies founded at more or less the same time in the early modern period, by more or less rapacious Europeans, with more or less the same intentions, North America became such an economic success, while Latin America did not. What explains “American exceptionalism”? There is no shortage of candidates, of course, but one of the most prominent is the notion of the “Frontier”.¹ Many scholars have claimed that a crucial aspect of the uniqueness of the United States was the vastness of the open spaces (at least where the indigenous peoples had died, Mann (2005)) which heavily influenced the way society, economy and polity evolved.

The most famous exposition of this view, first developed in 1893, was due to Frederick Jackson Turner. Turner, developing what has become known as the “Frontier (or Turner) thesis”, argued that the availability of the frontier had led to a particular type of person and had crucially determined the path of US society.

“The existence of an area of free land, its continuous recession, and the advance of American settlement westward, explain American development”.

“Behind institutions, behind constitutional forms and modifications, lie the vital forces that call these organs into life and shape them to meet changing conditions” Turner (1920, pp. 1-2).

Turner emphasized that the frontier created strong individualism and social mobility and his most forthright claim is that it was critical to the development of democracy. He noted

“the most important effect of the frontier has been to promote democracy” Turner (1920, p. 30).

and

“These free lands promoted individualism, economic equality, freedom to rise, democracy... American democracy is fundamentally the outcome of the experiences of the American people in dealing with the West” Turner (1920, pp. 259, 266).

Moreover, the things that went along with democracy and helped to promote it, such as social mobility, most likely also stimulated economic performance.

Even if some have deferred since Turner wrote, the “Frontier Thesis” has become part of the conventional wisdom amongst historians and scholars of the United States². Though the specific mechanisms that Turner favored, such as individualism, have become less prominent, arguments about the frontier have appeared in many places, particularly the literature on the democratization of the United States (Keyssar (2000), Engerman and Sokoloff (2005)). Keyssar (2000, p. xxi) argues

“The expansion of suffrage in the United States was generated by a number of key forces and factors ... These include the dynamics of frontier settlement (as Frederick Jackson Turner pointed out a century ago)”.  

¹For other ideas on this topic see Hartz (1955), Hartz (1964) and Lipset (1996).
²For some of the debate about the applicability of this thesis to the United States see Taylor (1956), Billington (1966), Hofstadter and Lipset (1968).
Those who have contested this view have tended to focus on the extent to which the Frontier did or did not have the postulated effects within the United States.

At some level the acceptance of the Frontier thesis and the nature of the debate is quite surprising. This is because the existence of a frontier clearly did not distinguish the United States from the other colonies of the Americas or indeed other societies such as Russia, South Africa or Australia in the 19th century. Every independent South American and Caribbean country, with the exception of Haiti, had a frontier in the 19th century. As in the United States, these frontiers were usually inhabited by indigenous peoples and they went through the same pattern of expansion into this zone which, as in the United States, coincided with the expropriation and oftentimes annihilation of indigenous communities. In these cases, however, there seems to be much less reason to associate the frontier expansion with democracy or economic development. Indeed, one could conjecture that if the Frontier thesis had been developed by Latin American academics in the late 19th century it would have been formulated with a minus sign in front.

A small literature has examined the frontier hypothesis in comparative perspective, but it has come to inconclusive results. Turner did engage in some comparative observations but refers only to Europe, noting

“...The American frontier is sharply distinguished from the European frontier - a fortified boundary line running through dense populations Turner (1920, p. 3).

Hennessy (1978) specifically addresses the applicability of the Frontier thesis to Latin America (see also the papers in Weber and Rausch (1994)). Noting the absence of a literature on the Frontier thesis in Latin America Hennessy (1978, p. 13) reasons

“If the importance of the Turner thesis lies in its ... ability to provide a legitimating and fructifying nationalist ideology, then the absence of a Latin American frontier myth is easy to explain. Without democracy, there was no compulsion to elaborate a supportive ideology based on frontier experiences”.

Hennessy’s general conclusion is that the thesis is irrelevant because

Latin American frontiers have not provided fertile ground for democracy. The concentration of wealth and the absence of capital and of highly motivated pioneers effectively blocked the growth of independent smallholders and a rural middle class Hennessy (1978, p. 129).

The correlation between good outcomes and the frontier in the United States and Canada but the lack of such a correlation in Latin America raises the question of whether or not in general there is any connection between the frontier and economic and political development. Maybe the frontier was irrelevant? A myth? We believe the answer to this is no. Some of the mechanisms described in the case of the United States certainly seem plausible, it is just that they don’t seem to have operated in Latin America. The key to understanding why comes from examining how frontier land was allocated. In the United States it was

---

3 Though the issue of the role of the frontier has been considered in Latin America studies, see Hennessy (1978) and Weber and Rausch (1994), it appears that nobody has made these comparative observations before.

4 Other work looking, usually critically, at the Frontier thesis is comparative perspective include Winks (1971), Miller (1977), and Powell (1981). For more general discussions of frontier expansions in the modern world not focused on the Turner thesis see Richards (2000) and Belich (2009).

5 Differences in labor institutions developed in frontier areas possibly also might played a role.
the 1862 Homestead Act which played the major role in governing who and on what terms had access to
the frontier. In Latin America, on the other hand, only Costa Rica and Colombia passed legislation which
resembled this in practice. In a few other countries where some legislation was passed, it seems to have never
been put in practice. Jeffer (1926, p. 167), for example, points out the difference between the "elevated
aims and philanthropic language" of the Argentine legislation regarding landowning in frontier areas and "the
actuality of events". More generally, frontier land was allocated in a very oligarchic pattern by existing elites,
and property rights over frontier lands of settlers were in many cases weak. Though Turner continually talks
about the frontier and ‘free land’ as if they were the same thing, as Adelman (1994, p. 101) points out

"Turner... overlooked two hard facts: land was not free, and workers had to be brought in from
outside the region.

Outside of Costa Rica and Colombia, frontier land was not free in Latin America and indeed was allocated
oligarchically by those with political power. Hennessy (1978, p. 19) observed

"Another contrast lies in the availability of ‘free land’. Whereas free land was the magnet attracting
pioneers into the North American wilderness, in Latin America most available land had been
preempted by landowning patterns set in the sixteenth century.

The historical experience of Argentina is again revealing. Jefferson (1926, p. 175-178) describes several
episodes in the Paraná basin, the Nequén region to the South or even in La Pampa, where settlers found
difficulties in maintaining their property rights over the lands they opened, both because State officials
reigned on past promises or because of abuses from local elites. Interestingly, when Turner does discuss
the issue of land laws with respect to the frontier, he seems to see these as an exogenous response to the
existence of the frontier, for example arguing that

"The disposition of the public lands was a third important subject of national legislation influenced
by the frontier" Turner (1920, p. 25).

and

"It is safe to say that the legislation with regard to land ... was conditioned on frontier ideas and
needs" Turner (1920, p. 27).

These arguments suggest to us not that the frontier is irrelevant, but rather that a more nuanced version
of the Frontier thesis is required. We refer to this as the “conditional Frontier thesis”. This takes into
account the fact that the consequences of the frontier are conditional on the initial political equilibrium when
frontier expansion occurred. Although the opening up of a frontier might bring new opportunities for the
establishment of equitable societies as Turner suggested, in relatively oligarchic countries the existence of
an open frontier gave the ruling elite a new valuable instrument which they could manipulate to remain in

---

There is a large historical literature on the oligarchic allocation of frontier lands in 19th century Latin America. For
overviews of the central American experience see Williams (1994), Gudmundson (1997) and Mahoney (2001); McCreery (1976)
and McCreery (1994) for the Guatemalan experience. Parsons (1949) is the classic work on frontier expansion in Colombia, see
also Christie (1978) and LeGrand (1986). Dean (1971) and Butland (1966) analyze the Brazilian case. Solberg (1969) presents
Argentina and both books make interesting comparisons to the differential evolution of Canada.
power. They did this through the structure of land and laws, policies towards immigrants and clientelistic access to frontier lands. When initial political institutions were different, as they were in the United States, Canada, Costa Rica and Colombia, elites were less able to manipulate this resource and a more open society evolved. As Turner argued, it is quite likely in these circumstances that the existence of a frontier helped to induce further improvements in political institutions. In countries like Argentina or Mexico, it is possible that an oligarchically allocated frontier was worse than having no frontier at all.

In this paper we propose what we believe is the first empirical test of the Frontier thesis and particularly the “conditional Frontier thesis”. To do this we construct an estimate of the proportion of land which was frontier in each independent country in the Americas in 1850. We combine this with data on current income per-capita, democracy and inequality. Our first main finding is that our estimates of the relative size of the frontier are positively correlated with long-run economic growth and the extent to which countries were democratic over the 20th century. The relative size of the frontier is also negatively correlated with income inequality. These initial results are quite consistent with the simple Frontier thesis.

Nevertheless, we then test for the “conditional Frontier thesis” by interacting the proportion of frontier land in 1850 with measures of initial institutions, specifically constraints on the executive from the Polity dataset which is available for every independent country in the Americas in 18507. When GDP per-capita in 2007 is the dependent variable we find that neither frontier land in 1850 nor constraints on the executive are themselves statistically significant, but their interaction is. Indeed, the results imply that for countries with the lowest level of constraints on the executive (which is almost half our sample in 1850) long-run economic growth is lower the larger is the frontier. For higher levels of constraints, however, long-run growth is higher. These simple regressions are very consistent with our hypothesis. With respect to democracy, when we look at the average Polity Score from 1900-2007 we again find that once we add the interaction term, neither frontier nor constraints themselves are significant. In this case we do not find that the frontier is ever bad for democracy, but rather its impact on democracy is greater the greater are constraints on the executive in 1850. These results suggest, consistent with the “conditional Frontier thesis”, that the frontier on its own had no impact on democracy. When we turn to the democracy score averaged over the post World War II period (1950-2007) we find different results. Here frontier on its own tends to be positively correlated with democracy while the interaction term is not statistically significant. Finally, when we examine contemporary inequality as the dependent variable we do not find robust results. Though frontier and constraints on the executive in 1850 are both negatively correlated with inequality, when we add the interaction term none of the variables is statistically significant.

Our argument about the conditional effect of the frontier is related to several important historical debates. For example, one interpretation of the arguments of Brenner (1976) is that large shocks in the middle ages, such as trade expansion or the Black Death had conditional effects which depended on initial institutions. In Britain where the serfs were relatively organized and where Lords did not have large estates, the Black Death empowered the lower orders and led to the collapse of feudal institutions. In eastern Europe, however, where the initial conditions were different, the Black Death ultimately led to the “Second Serfdom”. A related argument is presented in Acemoglu et al. (2005) who argue that the impact on Western Europe of trade and colonial expansion after 1492 depended on initial political institutions. In places where there were relatively strong political institutions, such as Britain and the Netherlands, trade expansion led to improvements of institutions and stimulated economic growth and further political change. In places which

7 Except for Canada, for which data is available starting in 1867.
were more absolutist, such as Spain and France, trade expansion had opposite effects.

The paper proceeds as follows. In the next section we discuss how we measure the extent of the frontier and present some basic data about its extent and nature. In section 3 we examine the correlation between the frontier and long run economic and political outcomes. Section 4 investigates whether or not there is a conditional effect of the frontier and section 5 concludes.

2 Measuring the Frontier

The literature on the frontier has been quite vague on how exactly to determine what was or what was not frontier. Turner himself noted (1920, p. 3)

"In the census reports it is treated as the margin of that settlement which has a density of two or more to the square mile. The term is an elastic one, and for our purposes does not need a sharp definition. We shall consider the whole frontier belt, including the Indian country and the other outer margin of the "settled area" of the census reports.

It was the definition of the frontier as areas with a population density of less than two people per square mile that led the census bureau to declare in 1890 that the US frontier had closed.

Any attempt to measure the extent of the frontier across the Americas must confront several methodological issues. In the first place, frontiers in each country, and even within each country, looked very different around the mid-nineteenth century. Coming up with a measure of the frontier for each country therefore requires a compromise to select some basic simplifying but consistent criteria which will necessarily overlook many possibly important dimensions. Following the historical literature and the relevance of the criteria in the light of Turner's arguments regarding the frontier, the natural candidates for such a classification are the presence or absence of native American communities not subject to state control and authority, overall population density (including any non-native American settlers), and the presence or absence of state institutions. All of these conditions were important determinants of the potential availability of free land and of the possibilities for successful settlement. Especially problematic is that we would like to think of the frontier as a dichotomous condition, whereas its defining variables are in most cases inherently continuous, and its boundaries usually not clear-cut.

When dealing with the frontier experience of South America another issue arises; settlement of frontier lands was not an absorbing state in some regions. Several areas in Paraguay, for example, were significantly settled and run by Jesuit missionaries during the colonial period. After the expulsion of Jesuits from the Spanish Empire in 1767, the Crown reassigned the control of these regions to other religious communities who failed in maintaining the economic success of the missions and the political control of the indigenous communities inhabiting the areas. As a result, in a matter of decades the missionary regions degenerated to a virtual absence of State control and became frontiers once again. They remained as such until late in the 19th century (Eidt (1971), Moniz (2006)). The case of Brazilian bandeirantes in the 17th and 18th centuries is similar. Brazil expanded its boundaries as these settlers moved west into the Amazon and its south-western basin. Nonetheless, many of these areas were subsequently unsettled and remained like that until late in the republican period. As a result, Brazilian historiography refers to them as "hollow frontiers" (Katzman

\*This type of interaction also comes up in the literature of the impact of the resource curse, see Moene et al. (2006).

170
For our purposes we tried to include in our measure these regions, which around 1850 were in fact not controlled by republican states even if they had been so earlier in colonial times.

Once such decisions have been made, the second issue is related to the availability of information about the definitional criteria for what frontier and non-frontier lands were. Not only is detailed information scarce by the very nature of the subject, but the comparability of the data across countries might also be problematic. We collected three types of information, based on which we constructed three alternative measures of the frontier: (a) historical cartographic data depicting directly information on frontier territories or on population density for several of the countries in our sample of independent republics, at different dates starting in the mid 19th century, b) geographic (and georeferenced) information on current-day administrative divisions (provinces, departments or states), and c) direct country or regional historical accounts on the settlement of frontier areas during the 19th century. The appendix contains a detailed description of the sources used for each country. The reason that making use of current administrative divisions is helpful is that in fact the formation of administrative units in many regions across the Americas was precisely driven by significant settlement and State presence. The best examples of this might be the straight lines marking the boundaries of the western states of the United States, put in place as a first effort to regulate and control the newly occupied territories as the westward expansion moved on, or the Amazon rainforest frontier provinces of countries like Colombia, Brazil or Peru, which were designed precisely to delimit such frontier areas.

2.1 The Frontier in the United States and Canada

For these two countries we were able to find detailed cartographic information which allowed us to calculate the share of unsettled and settled land in 1850. More specifically, for the United States the Office (1898) and Gerlach (1970) contain detailed maps of population density. Both sources use the 19th century United States Census data, and following the Census Bureau, classify as frontier land the territory with less than 2 people per square mile (0.7725 people per square kilometer). For Canada, the of Statistics (n.d) contains maps for several years in the second half of the 19th century, depicting population density by points on the map. We directly georeferenced these maps using GIS software, and computed the share of total land area of each country with population density below 0.7725 people per square kilometer, in 1850 for the United States and in 1851 for Canada. Since these maps were based on detailed census data, we believe these frontier measures have the smallest possible measurement error, and are the only ones we consider for these two countries.

For the rest of countries in the Americas the information is not as detailed and is more scattered throughout different sources. As a result, we decided to create a set of alternative measures of the frontier, taking into account the differences we found when comparing the available information.

2.2 The Frontier in Central America

To measure the Frontier in Central America we relied heavily in Hall and PérezBrignoli (2003), which contains rich historical maps for Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica and Panama, of settlement during the 19th century, and also has a thorough historical discussion of the frontier expansion throughout the region. We merged the information of these maps, which depict the frontier regions in each country, with a georeferenced sub-national level map of Central America, and coded each province/department/State as frontier or non-frontier depending on whether or not it fell into the regions considered as unsettled in the
Hall and Pérez Brignoli (2003) maps. Of course, with this procedure a considerable number of sub-national units appeared as partially frontier areas. We thus created two different measures of the frontier, which we call narrow and wide. The narrow measure classifies as non-frontier the sub-national units for which an ambiguous coverage of the Hall and Pérez Brignoli (2003) maps had been obtained, while the wide measure classifies them as a frontier. We further refined the classification of provinces using of the Census (1956a), which contains very detailed population density maps for all the Central American republics in 1950 at the province/department level. The comparison with these maps allowed us to reclassify provinces that might have been ambiguous, but which by 1950 clearly had a population density below 0.7725 people per square kilometer, and necessarily must have been frontier areas 100 years before. The Appendix presents the coding of each sub-national unit in its narrow and wide versions.

For the Mexican frontier we relied on the of Business Research (1975) population density map for 1900, a state-level map based on the 1900 Censo General de Población, together with Bernstein (1964) and Hennessy (1978). Since population density in 1900 was considerably higher than in 1850 everywhere in Mexico, we coded as frontier states not only those with less than 0.7725 people per square kilometer in 1900, but also any State with at most a population density of 5 people per square kilometer in 1900, which were at the same time mentioned in the complementary references as frontier areas. This resulted in a relatively straightforward classification except for the state of Chiapas, which we coded as non-frontier in the narrow measure and as frontier in the wide measure.

2.3 The Frontier in the Caribbean Republics

Only Haiti and the Dominican Republic were independent by 1850, and as such are the only two Caribbean countries in our sample. Coding the frontier for them was a pretty straightforward job based on Anglade (1982) and Lora (2002). Anglade presents population density maps for the late 18th century, and mid 19th century, where it is clear that since the colonial period Haiti had population densities well above 0.7725 people per square kilometer, and almost everywhere significantly higher. Haiti therefore did not have a frontier. For the Dominican Republic the picture is very similar, except possibly for the provinces of Barahona and Pedernales in the south-western tip of the country. The of the Census (1956b) also contains detailed province-level maps of these two countries in 1950, which show a low population density in the southwest of the Dominican Republic. As a result, the narrow measure considers Barahona and Pedernales as non-frontier, while the wide measure codes them as frontier. All the rest of the country is coded as non-frontier.

2.4 The Frontier in South America

To measure the frontier in the South American countries we followed a procedure very similar to the one we used for the Central American republics, merging the information in usually country-specific historical maps and accounts with current-day sub-national units. The Appendix contains the historical references used for each country. When a sub-national unit was partially covered by settlement we again made the distinction by coding it as non-frontier in the narrow measure and as frontier in the wide version. This is the case, for example, of the north-eastern Brazilian province of Piauí or the Pacific coast province of Esmeraldas in Ecuador.

For South America we found an alternative source for the frontier. Butland (1966), which discusses in detail the frontier expansion in southern Brazil, presents a South American map depicting the frontier areas in mid
19th century. Unfortunately he does not explain how this map was drawn, but actually it coincides to a quite large extent with our own province-level codings. We used GIS software to georeference the frontier map in Butland (1966) and directly computed the share of each country which was frontier in the mid-19th century. As a result we have three different frontier measures for South America: narrow, wide and Butland.

Table 1 sums up the data from these calculations. For the United States and Canada we only have one number each, with 72.5% of the territory of the United States being frontier in 1850, while the corresponding number for Canada is 85.3%. Map 1 shows exactly where the frontier and non-frontier areas were. This is a pretty familiar picture with, for example, the United States being settled on the eastern seaboard and all the way east to the western boundaries of Arkansas and Missouri. Far to the west parts of coastal California and the central valley north of San Francisco were also settled. For the countries in South America we have three different estimates of the extent of the frontier. For example, Table 1 shows that for Colombia the narrow definition of the frontier suggests that 62.9% of the territory was frontier in 1850 and this exactly coincides with the wide definition. Butland's map gives a fairly similar estimate of 58.1%. For other countries, however, the differences between these estimates are much larger. For example, for Argentina the narrow definition is 49.3% while the wide one is 74.2%. The reason for this large difference is easy to see from Map 2. Here the settled areas intersect with many departments. For instance the narrow definition treats the departments of San Luis, Córdoba, Neuquén, Santiago del Estero and Salta as settled, while the wide definition treats them as frontier. For Argentina, Butland's estimate is close to our wide definition. Finally, Map 3 looks at Central America and the Caribbean.

These calculations clearly illustrate our conjecture from the introduction which is that simply in terms of the size of the frontier, the United States is not distinct. Uruguay had a frontier which was quite a bit larger relative to the size of the country and Brazil's frontier was also larger. Other countries such as Costa Rica, Nicaragua or Venezuela had frontier's which were only about 15% or so less.

3 Other Data

Apart from the data we constructed on the extent of the frontier in 1850, we use some other readily obtainable data. For our measure of historical political institutions we use constraints on the executive in 1850 from the Polity IV Project. This variable is defined as the extent of institutional restrictions on decision making powers of the chief executive, whether individual or collective. In a democracy constraints would come from the legislative or judicial branches of government. In a dictatorship constraints may come from the ruling party in a one-party system, a council of nobles or powerful advisors in monarchies, or maybe the military in polities which are subject to the threat of military coups. The extent of constraints on the executive are coded as being between 1, meaning “unlimited executive authority” and 7, implying “executive parity or subordination”. A country would be in the first category if “constitutional restrictions on executive action are ignored” or “there is no legislative assembly or there is one but it is called or dismissed at the executive’s pleasure”. A country would be in the latter category if “a legislature, ruling party or council of nobles initiates much or most important legislation” or “the executive is chosen by the accountability group and is dependent on its continued support to remain in office.”

Figure 1 shows the distribution of constraints on the executive in 1850 for the 21 countries in our dataset. One can see that 9 countries are assigned the minimum score of 1, while the United States and Canada have

the maximum score of 7\(^{10}\). Interestingly for our hypothesis, Costa Rica and Colombia both have scores of 3 in 1850. The country with constraints of 5 in 1850 is Honduras.

We also use the Polity IV Project’s measure of how democratic a country is, which they refer to as the Polity IV score, which is the difference between the Polity’s Democracy and Autocracy indices\(^{11}\). The democracy score ranges from 0 to 10 and is derived from coding the competitiveness of political participation, the openness and competitiveness of executive recruitment and constraints on the chief executive. The Polity Autocracy Index also ranges from 0 to 10 and is constructed in a similar way to the democracy score based on scoring countries according to competitiveness of political participation, the regulation of participation, the openness and competitiveness of executive recruitment and constraints on the chief executive. This implies that the Polity IV score ranges from -10 to 10.

The other data we use is GDP per-capita in 2007 PPP adjusted from the World Bank’s World Development Indicators CD Rom and from the same source we also take information of the Gini coefficient for income distribution which we average over the period 1996-2005.

Table 2 shows some basic descriptive statistics of the data. The rows correspond to our different dependent and key explanatory variables and we divide the sample according to the median extent of frontier land in 1850 according to our narrow definition. The first set of columns show the average data for countries with greater than median frontier land, while the last set of columns in the table show the data for less than median frontier land. The median country here is Mexico, 57% of whose land was frontier in 1850 according to our narrow definition. Note that for countries below the median the average amount of land which was frontier was 32% (with a standard deviation of 0.22), while for countries above the median the average proportion of frontier land was 70% (with standard deviation of 0.12).

The comparison of low and high frontier countries is quite revealing. For instance looking at the third row of Table 2 we see that GDP per-capita in 2007 on average was $11,466 for above median frontier societies, while it was only $3,744 for below median. The data shows that those countries which had a relatively large frontier in 1850 now have substantially higher income per-capita. In row 4 we show the average Polity IV score over the period 1900-2007. This is 2.43 for above median countries and -0.35 for below median. In the next row we instead look at the average Polity IV score for the period 1950-2007. Though there is a clear upward trend in the extent of democracy, the comparison looks quite similar with above median frontier countries which have an average polity score of 3.96 while below median countries have a score of 1.05. As with income per-capita, there seems to be a clear pattern with countries which had relatively large frontiers in 1850 being today more democratic than those which had relatively small frontiers in 1850.

Finally, the last row examines average inequality over the period 1996-2005. The average Gini coefficient for high frontier countries is 49.1 while for low frontier countries it is 53.4. Just as countries with relatively large frontiers are more prosperous and democratic, they also appear to be more equal.

These raw numbers are quite consistent with the basic Frontier thesis. It is interesting to examine them in figures. Figure 2 plots the share of frontier (narrow definition) against GDP per-capita in 2007. There is a pronounced positively sloped relationship which remains even if the United States and Canada are dropped. Figure 3 examines the raw relationship between the share of frontier land against the Polity score over the

\(^{10}\) As previously noted, Polity data for Canada only starts in 1867, at which point it has a 7, which we used as the its 1850 number.

\(^{11}\) This measure is a very standard one in empirical work on democracy, and other definitions typically give very similar results (see Acemoglu et al. (2008)).
period 1900-2007. The picture is rather similar with a distinct positive correlation and with North America and Costa Rica far off the regression line. Figure 4 shows the same picture but now with the Polity IV score averaged over the post World War II period, 1950-2007. This is very similar to Figure 4. Finally, Figure 5 examines inequality and the extent of the frontier. This Figure suggests that there is a negative correlation between the extent of the frontier and contemporary inequality.

All of the above give support to the Turner Thesis. We now turn to regression analysis to investigate how robust they are and whether these numbers may also be consistent with our conditional Frontier thesis. As we shall see, the image which emerges from the descriptive statistics and simple scatterplots is not general.

4 Empirical Results

We now examine some simple regression models to examine the long-run consequences for economic and political development of having a frontier. In all cases we estimate Ordinary Least Squares regressions of the form

\[ y_i = \alpha + \beta F_{i,1850} + \gamma C_{i,1850} + \delta (F_{i,1850} \times C_{i,1850}) + \varepsilon_i \]  

(1)

where \( y_i \) is the dependent variable of interest for country \( i \). This is respectively GDP per-capita in 2007, the democracy score of Polity averaged over different periods, or the Gini coefficient of inequality averaged over some period. \( F_{i,1850} \) is the proportion of the country which was frontier land around 1850, \( C_{i,1850} \) is constraints on the executive from Polity in 1850, and \( \varepsilon_i \) is a disturbance term which we assume to have the usual properties. Here, following the discussion above, we also allow for the interaction between constraints on the executive and frontier land in 1850.

4.1 Income Per-Capita

We first look at regressions where \( y_i \) is GDP per-capita for country \( i \) in 2007. These are recorded in Table 3. The table is split into three sets of columns where each set uses a different definition of the frontier. The first three columns use our narrow definition of the frontier, the second three our wide definition and the final three columns use the Butland definition\(^{12}\).

The first column shows the most parsimonious OLS regression of GDP per-capita on the proportion of land that was frontier in 1850. The coefficient \( \beta = 18,324.1 \) (with a standard error of 9,953.3) is statistically significant. To see what this coefficient implies, consider Mexico, which is the median frontier country, with 57% of its territory comprised of frontier. This coefficient implies a GDP per-capita for Mexico of \( -1,738 + 18,324 \times 0.57 = 8,706 \), which is pretty close to the actual value for Mexico which is 8,340. The coefficient on the frontier share implies that if one changed the frontier from the median level to the level of the United States, which is 0.72, GDP per-capita would increase by \( (0.72 - 0.57) \times 18,324 = 2,748 \), which is a 31% (= 2.748/8.706) increase of the predicted income for the median country. Alternatively, if Mexico’s frontier increased by 10%, from 57% to 62.7%, income would increase by \( (0.627 - 0.57) \times 18,324 = $1,044.5 \).

It is important to note, however, that one should be very cautious about proposing any type of causal interpretation of the data. For example, we have treated the extent of the frontier in 1850 as econometrically

\[^{12}\text{Since the Butland data are only available for the South American countries, the Butland frontier definition uses the narrow frontier measure for the rest of the sample.}\]
exogenous, while in fact it may be the endogenous outcome of other factors that influence economic or political development. Perhaps countries that had good fundamentals had expanded more, for instance by attracting greater numbers of migrants, and thus tended to have relatively small frontiers in 1850. Of course if this form of omitted variable bias were important, it actually suggests that we might be underestimating the effect of the frontier because it suggests that relatively small frontiers ought to be associated with factors that also lead to good long-run development. We are also treating constraints on the executive as exogenous, which is again unlikely to be the case.

In column 2 we add constraints on the executive in 1850. This greatly increases the extent of variation explained by the model and both constraints and frontier are significant, though the estimated coefficient on frontier falls. The coefficient on constraints, $\gamma = 4,405.86 \ (s.e. = 1,346.5)$ is statistically significant.

Column 3 then adds the interaction term. This term is highly significant, $\delta = 11,843.7 \ (s.e. = 3,015.5)$ and the estimated coefficient on frontier now changes sign so that $\beta = -13,489.29 \ (s.e. = 7,835.69)$. One can see here that when constraints on the executive are equal to 1 (which is the case in 9 out of our 21 countries in 1850) the total effect of frontier is $\beta + \delta \times 1 = -13,489.29 + 11,843.7 = -1,645.59 < 0$. In other words for countries with the lowest value of constraints on the executive, representing “unlimited executive authority” the greater is the relative size of the frontier in 1850, the poorer is the country today. However, as long as constraints are 2 or above, frontier land is positively correlated with long-run growth.

It is also interesting to examine the quantitative impact of these results. For example, if we held the extent of frontier fixed and increased the level of constraints on the executive in a country from 1 to 7, then this would imply a change in income of

$$(-13,849 \times F_{1850}) + (11,843 \times F_{1850} \times 6)$$

$$= (-13849 \times F_{1850}) + (71,058 \times F_{1850}) - 21,942 = (57,209 \times F_{1850}) - 21,942$$

Hence, a country with median frontier would increase its current income by $0.57 \times 57,209 - 21,942 = 10,667$ which would eliminate about one third of the income gap between Mexico and the United States.

Columns 4-6 then re-estimate the same 3 models using our wide definition of the frontier. The results are very similar to those in the first three columns with the narrow definition except that now neither frontier nor constraints on the executive are significant when they are entered with the interaction. The final three columns use the Butland definition of the frontier with similar results.

In all specifications when we enter the interaction term it is robustly estimated and very significant and in all cases suggests that when constraints are at their minimum, the presence of the frontier was bad for economic development, while at higher levels of constraints, the frontier was good for long-run economic growth. The results in this section are not consistent with the Frontier thesis but they are consistent with the conditional Frontier thesis.

4.2 Democracy

We now turn to regressions where $y_i$ is the Polity score for country $i$ averaged over different periods. We look at two such periods, one is 1900-2007 and the other is 1950-2007. These regressions are in Tables 4 and
respectively. As with Table 3, each table is split into three sets of columns where each set uses a different definition of the frontier.

Table 4 column 1 shows the simplest regression of the Polity score 1900-2007 on frontier in 1850. There is a significant positive correlation with $\beta = 8.189$ (s.e. = 2.458). The second column adds constraints on the executive in 1850. Constraints are also significantly positively correlated with democracy in the 20th century with an estimated coefficient of 1.474 (s.e. = 0.195).

The third column then adds our interaction term. The interaction term is marginally significant with a t-statistic of 1.78 and has a positive coefficient of $\delta = 1.263$. However, unlike in the regressions where income per-capita was the dependent variable, the frontier share on its own remains positive and significant, even if the magnitude of the coefficient falls by 50%.

The rest of Table 4 shows that these results are not completely robust. The interaction terms remains positive and basically significant, but when we use the wide definition of the frontier, frontier entered on its own is not statistically significant in column 6, or using the Butland definition in column 9. Nevertheless, there is no evidence here of any negative effect of the frontier, unlike in the income regressions. The results in Table 4 suggest that even for the lowest level of constraints on the executive, the greater was the frontier in 1850, the more democratic the country was in the 20th century. Nevertheless, the quantitative effect is larger, the greater are constraints in 1850.

In Table 5 we re-estimate the same models as in Table 4 except that now we average the dependent variable only over the post World War II period. As is quickly seen this gives some quite different results. When we just control for frontier and constraints on the executive, the results in terms of the size and significance of the coefficients are very similar to those in Table 4. However, once we control for the interaction we find that the interaction term is never close to significant while the estimated coefficient on frontier on its own remains more or less the same quantitatively and mostly significant (only marginally so in column 6). This table shows that the conditional effect on democracy is actually a phenomenon of the first half of the 20th century. In the second half the simpler version of the Frontier thesis captures the patterns in the data quite nicely.

4.3 Inequality

Finally, we let $y_i$ in equation (1) be the average Gini coefficient for country $i$ over the period 1990-2007. The results of estimating this model are reported in Table 6. A quite robust pattern emerges in all three sets of columns, irrespective of how we measure the extent of the frontier. When entered on its own, frontier is negatively and significantly correlated with contemporary income inequality, as are constraints on the executive. These results suggest that either having a bigger frontier in 1850 or better political institutions is associated with lower inequality today. However, as columns 3, 6 and 9 indicate, once the interaction term is included none of the coefficients are statistically significant.

5 Conclusions

In this paper we have developed what to our knowledge is the first test of the “Frontier (or Turner) thesis”. Turner argued that it was the existence of the frontier that generated the particular path of development that
the United States followed in the 19th century. Though his work on the United States has been criticized, it still appears to heavily influence the ways scholars think about these issues. The basis of our assessment has been the observation that every country in the Americas, with the possible exception of El Salvador and Haiti, had a frontier in the 19th century. The United States was certainly not exceptional in either this or the relative extent of the frontier. In consequence, seen in comparative context, the existence of a frontier does not seem to be obviously correlated with long run economic and political development.

We hypothesized, however, that there may be a conditional relationship between the extent of the frontier and political institutions at the time of the allocation of frontier land. Historical evidence suggests that even if most countries in the Americas had an open frontier, how that frontier land was allocated differed a lot. For example, while the United States, Costa Rica and Colombia passed Homestead Acts or something approximating them, in places like Argentina, Chile or Guatemala, political elites allocated frontier lands to themselves or associates in a very oligarchic manner. This suggests that the impact of the frontier might be conditional on the existing political institutions which influenced how the land was allocated - a notion we dubbed the “conditional Frontier thesis”. Our hypothesis suggests that if political institutions were bad at the time of frontier settlement, the existence of such frontier land might actually lead to worse development outcomes, probably because it provides a resource which non-democratic political elites can use to cement themselves in power.

To investigate more systematically the relationship between the frontier and long-run development we constructed measures of the extent of frontier land for 21 independent countries in the Americas in 1850. Using some simple regressions we showed that the data does indeed support our conditional hypothesis. With respect to both income per-capita today and democracy over the 20th century, it is the interaction between the extent of the frontier in 1850 and constraints on the executive in 1850 that plays the primary explanatory role. For example, for a country with the lowest level of constraints on the executive, the larger is the relative size of the frontier, the lower is GDP per-capita today. For countries with higher constraints, however, a larger frontier is positively correlated with current GDP per-capita. With respect to democracy we found that for a given level of constraints in 1850, greater size of the frontier is correlated with greater democracy in the 20th century, though this effect comes primarily from the first half of the century.

There are many caveats with these findings. For example, we did not control for variation in the “quality” of the frontier. For instance there may be a big difference between Oklahoma in the United States and the Atacama Desert in northern Chile, both of which were frontiers in 1850. Still, the United States also had large areas of the Rocky Mountains which were not high quality lands. Trying to control or adjust for this explicitly is an important area for future research. We also intend to conduct more sensitivity analysis. While 1850 seemed to us to be an interesting year to focus on because it marked the beginning of the period of the rapid expansion of world trade which created such huge frontier movements in the Americas, one could argue it is too late. An important area for future research is a more intensive sensitivity analysis than is presented here.

Nevertheless, results suggest that the role of the frontier is much more complex than the Turner thesis suggests. The consequences of the existence of a frontier for different countries in the Americas depended a lot on the nature of political institutions which formed in the early independence period. If these institutions featured few constraints on the executive, having a frontier was actually bad for economic development. If El Salvador and Haiti had had frontiers in the 19th century, this would have made them poorer today, not richer. Though we found no such negative effect for democracy, we did find that the impact of the frontier
on the democratization of a society was conditional on initial political institutions. If Turner thought that the United States frontier had a strong democratizing effect, this was only because it was in a country which already had good political institutions. This effect was severely muted in Latin America.

Though our results are not consistent with a large part of the Turner thesis, they are consistent with the research of Brenner (1976) and Acemoglu et al. (2005) which emphasized that the implications of large shocks or new economic opportunities depends on the initial institutional equilibrium. More specifically in the Americas, they are also consistent with the work of Engerman and Sokoloff (1997) and Acemoglu et al. (2001) and Acemoglu et al. (2002) who emphasized the critical importance of the creation of institutions in the colonial period and their path dependent consequences. In a sense, our results on income per-capita show how different paths were reinforced by the availability of frontier lands in the 19th century.
References


Dueñas, Carmen, Historia Económica y Social del Norte de Manabí, Quito: Abya Yala, 1986.


### Table 1

<table>
<thead>
<tr>
<th>Country</th>
<th>Total Number of Subnational Units</th>
<th>Total Land Area (square Kms.)</th>
<th>Number of Narrow Frontier Subnational Units</th>
<th>Total/Narrow Frontier Land Area (square Kms.)</th>
<th>Number of Wide Frontier Subnational Units</th>
<th>Total/Wide Frontier Land Area (square Kms.)</th>
<th>Total Frontier from Butland (1966) and Historical Cartography</th>
<th>Frontier Share from Butland (1966) and Historical Cartography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>24</td>
<td>2,160,403</td>
<td>11</td>
<td>1,370,454</td>
<td>15</td>
<td>2,068,542</td>
<td>74.2%</td>
<td>192,231</td>
</tr>
<tr>
<td>Bolivia</td>
<td>9</td>
<td>1,096,581</td>
<td>4</td>
<td>685,635</td>
<td>4</td>
<td>803,853</td>
<td>73.2%</td>
<td>861,507</td>
</tr>
<tr>
<td>Brazil</td>
<td>27</td>
<td>8,498,321</td>
<td>15</td>
<td>6,354,737</td>
<td>17</td>
<td>7,102,001</td>
<td>84.6%</td>
<td>7,026,006</td>
</tr>
<tr>
<td>Chile</td>
<td>13</td>
<td>756,095</td>
<td>5</td>
<td>398,745</td>
<td>5</td>
<td>398,745</td>
<td>52.7%</td>
<td>562,762</td>
</tr>
<tr>
<td>Colombia</td>
<td>33</td>
<td>1,141,748</td>
<td>15</td>
<td>718,130</td>
<td>15</td>
<td>718,130</td>
<td>62.0%</td>
<td>663,584</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>7</td>
<td>51,102</td>
<td>4</td>
<td>32,870</td>
<td>5</td>
<td>43,011</td>
<td>84.2%</td>
<td>32,870</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>32</td>
<td>46,891</td>
<td>0</td>
<td>-</td>
<td>2</td>
<td>3,665</td>
<td>7.8%</td>
<td>32,870</td>
</tr>
<tr>
<td>Ecuador</td>
<td>23</td>
<td>256,370</td>
<td>7</td>
<td>116,519</td>
<td>9</td>
<td>151,309</td>
<td>59.0%</td>
<td>120,827</td>
</tr>
<tr>
<td>El Salvador</td>
<td>14</td>
<td>21,040</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0.0%</td>
<td>-</td>
</tr>
<tr>
<td>Guatemala</td>
<td>22</td>
<td>108,889</td>
<td>2</td>
<td>44,892</td>
<td>7</td>
<td>59,692</td>
<td>64.0%</td>
<td>91,601</td>
</tr>
<tr>
<td>Honduras</td>
<td>18</td>
<td>112,492</td>
<td>3</td>
<td>45,262</td>
<td>6</td>
<td>64,904</td>
<td>57.3%</td>
<td>76,724</td>
</tr>
<tr>
<td>Haiti</td>
<td>9</td>
<td>27,700</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0.0%</td>
<td>-</td>
</tr>
<tr>
<td>Mexico</td>
<td>32</td>
<td>1,970,774</td>
<td>11</td>
<td>1,131,090</td>
<td>12</td>
<td>1,207,619</td>
<td>61.3%</td>
<td>130,827</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>17</td>
<td>120,339</td>
<td>4</td>
<td>77,129</td>
<td>7</td>
<td>91,601</td>
<td>76.1%</td>
<td>130,827</td>
</tr>
<tr>
<td>Panama</td>
<td>12</td>
<td>75,071</td>
<td>6</td>
<td>35,102</td>
<td>7</td>
<td>46,773</td>
<td>62.3%</td>
<td>59,692</td>
</tr>
<tr>
<td>Peru</td>
<td>25</td>
<td>1,285,199</td>
<td>4</td>
<td>595,813</td>
<td>7</td>
<td>709,235</td>
<td>55.2%</td>
<td>59.692</td>
</tr>
<tr>
<td>Paraguay</td>
<td>18</td>
<td>408,752</td>
<td>3</td>
<td>249,925</td>
<td>13</td>
<td>378,370</td>
<td>93.0%</td>
<td>365,065</td>
</tr>
<tr>
<td>Uruguay</td>
<td>19</td>
<td>175,016</td>
<td>19</td>
<td>175,016</td>
<td>19</td>
<td>175,016</td>
<td>100.0%</td>
<td>175,016</td>
</tr>
<tr>
<td>Venezuela</td>
<td>25</td>
<td>916,440</td>
<td>6</td>
<td>598,945</td>
<td>8</td>
<td>707,231</td>
<td>77.2%</td>
<td>655,533</td>
</tr>
<tr>
<td>United States</td>
<td>51</td>
<td>9,372,587</td>
<td>6</td>
<td>5,890,545</td>
<td>8</td>
<td>7,072,231</td>
<td>77.2%</td>
<td>6,792,227</td>
</tr>
<tr>
<td>Canada</td>
<td>13</td>
<td>9,017,699</td>
<td>6</td>
<td>5,890,545</td>
<td>8</td>
<td>7,072,231</td>
<td>77.2%</td>
<td>6,792,227</td>
</tr>
</tbody>
</table>

Map 1

The Frontier in North America circa 1850 (current administrative boundaries)

- Settled Areas
- Frontier Areas
Map 2

The Frontier in South America circa 1850 (current administrative boundaries)

- Settled Areas
- Frontier Areas
Map 3

The Frontier in Central America circa 1850 (current administrative boundaries)

- Settled Areas
- Frontier Areas
Figure 1

Constraints on the Executive in 1850

Source: Polity IV dataset. Note: Data for Canada is for 1867, and data for Panama is the same as for Colombia since in 1850 the former was a department of the latter.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Countries with Frontier Share &gt;= Sample Median Frontier Share</th>
<th>Countries with Frontier Share &lt; Sample Median Frontier Share</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>Mean</td>
</tr>
<tr>
<td>Share of Frontier Land circa 1850</td>
<td>11</td>
<td>0.700</td>
</tr>
<tr>
<td>Constraints on the Executive 1850</td>
<td>11</td>
<td>2.636</td>
</tr>
<tr>
<td>Per Capita Income 2007</td>
<td>11</td>
<td>11,466.38</td>
</tr>
<tr>
<td>Polly Score average 1900-2007</td>
<td>11</td>
<td>2.427</td>
</tr>
<tr>
<td>Income Gini average 1996-2005</td>
<td>11</td>
<td>49.113</td>
</tr>
</tbody>
</table>

Note: The sample median country for Frontier Share is Mexico, with a frontier share of 0.574 (based on our preferred measure of frontier). For the years in which the Polly score records a political transition, we assign the average score of the years before and after the transition, and years in which the Polly score assigns interruption or irregular periods are excluded from the averages.
Figure 2

Share of Frontier Land circa 1850 vs. GDP Per Capita in 2007

Source: GDP Per Capita is from the World Bank World Development Indicators (2008).  
Note: Share of Frontier land is our preferred measure (narrow frontier definition).

Figure 3
Share of Frontier Land circa 1850 vs. Polity IV Score
(average 1900-2007)

Source: Center for Systemic Peace, Polity IV Project (2007).
Note: Share of frontier land is our preferred measure (narrow frontier definition).
Figure 4

Share of Frontier Land circa 1850 vs. Polity IV Score (average 1950-2007)

Source: Center for Systemic Peace, Polity IV Project (2007).
Note: Share of Frontier land is our preferred measure (narrow frontier definition).
Figure 5

Share of Frontier Land circa 1850 vs. Income Gini (average 1996-2005)

Source: Income Gini is from the World Bank World Development Indicators (2008).
Note: Share of Frontier land is our preferred measure (narrow frontier definition).
<table>
<thead>
<tr>
<th></th>
<th>Narrow Frontier</th>
<th></th>
<th>Wide Frontier</th>
<th></th>
<th>Butland Frontier</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontier Share</td>
<td>18324.10</td>
<td>15777.35</td>
<td>-13849.29</td>
<td>10335.46</td>
<td>10397.25</td>
<td>-12590.71</td>
</tr>
<tr>
<td></td>
<td>(9953.30)</td>
<td>(4900.72)</td>
<td>(7835.69)</td>
<td>(6043.12)</td>
<td>(3884.45)</td>
<td>(8253.17)</td>
</tr>
<tr>
<td>Constraints on the Executive 1850</td>
<td>4405.86</td>
<td>-3657.29</td>
<td>4579.16</td>
<td>-3029.61</td>
<td>4708.54</td>
<td>-2663.75</td>
</tr>
<tr>
<td></td>
<td>(1346.50)</td>
<td>(2228.71)</td>
<td>(1528.40)</td>
<td>(3360.24)</td>
<td>(1371.11)</td>
<td>(2332.80)</td>
</tr>
<tr>
<td>Constraints 1850 x Frontier Share</td>
<td>11843.70</td>
<td>10391.53</td>
<td>10341.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3015.50)</td>
<td>(3765.30)</td>
<td>(2880.38)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.162</td>
<td>0.631</td>
<td>0.773</td>
<td>0.061</td>
<td>0.571</td>
<td>0.655</td>
</tr>
<tr>
<td>No. Observations</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

Note: Robust Standard Errors in parenthesis.
All regressions include a constant (omitted).
Table 4

<table>
<thead>
<tr>
<th>Frontier Share</th>
<th>Narrow Frontier</th>
<th>Wide Frontier</th>
<th>Butland Frontier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontier Share</td>
<td>5.264</td>
<td>4.737</td>
<td>1.839</td>
</tr>
<tr>
<td></td>
<td>(1.667)</td>
<td>(0.908)</td>
<td>(1.451)</td>
</tr>
<tr>
<td>Constraints on the Executive 1850</td>
<td>0.912</td>
<td>0.123</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.385)</td>
<td>(0.173)</td>
</tr>
<tr>
<td>Constraints 1850 x Frontier Share</td>
<td>1.159</td>
<td>1.703</td>
<td>1.123</td>
</tr>
<tr>
<td></td>
<td>(0.501)</td>
<td>(0.723)</td>
<td>(0.509)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.259</td>
<td>0.650</td>
<td>0.677</td>
</tr>
<tr>
<td></td>
<td>(0.501)</td>
<td>(0.723)</td>
<td>(0.509)</td>
</tr>
</tbody>
</table>

Note: Robust Standard Errors in parenthesis.
All regressions include a constant (omitted).
The Democracy score for Panama is average over the 1903-2007 period.
### Table 5

**Dependent Variable: Polity IV Score, average 1950-2007**

<table>
<thead>
<tr>
<th>Frontier Share</th>
<th>Narrow Frontier</th>
<th></th>
<th></th>
<th>Wide Frontier</th>
<th></th>
<th></th>
<th>Butland Frontier</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.213</td>
<td>7.455</td>
<td>9.809</td>
<td>5.622</td>
<td>5.780</td>
<td>5.474</td>
<td>5.304</td>
<td>5.815</td>
<td>7.597</td>
</tr>
<tr>
<td></td>
<td>(2.960)</td>
<td>(1.851)</td>
<td>(2.676)</td>
<td>(3.119)</td>
<td>(2.151)</td>
<td>(4.386)</td>
<td>(2.873)</td>
<td>(1.865)</td>
<td>(3.866)</td>
</tr>
<tr>
<td>Constraints on the Executive 1850</td>
<td>1.313</td>
<td>1.954</td>
<td>1.394</td>
<td>1.624</td>
<td>1.448</td>
<td>1.980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.254)</td>
<td>(0.959)</td>
<td>(0.282)</td>
<td>(1.197)</td>
<td>(0.252)</td>
<td>(1.080)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constraints 1850 x Frontier Share</td>
<td>-0.941</td>
<td>-0.314</td>
<td>-0.747</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.202)</td>
<td>(1.514)</td>
<td>(1.541)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.262</td>
<td>0.599</td>
<td>0.066</td>
<td>0.150</td>
<td>0.533</td>
<td>0.533</td>
<td>0.134</td>
<td>0.545</td>
<td>0.550</td>
</tr>
<tr>
<td>No. Observations</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

*Note: Robust Standard Errors in parentheses.*

All regressions include a constant (omitted).
Table 6

<table>
<thead>
<tr>
<th></th>
<th>Narrow Frontier</th>
<th></th>
<th>Wide Frontier</th>
<th></th>
<th>Butland Frontier</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraints on the Executive 1850</td>
<td>-1.740</td>
<td>0.117</td>
<td>-1.845</td>
<td>-0.147</td>
<td>-1.906</td>
<td>0.580</td>
</tr>
<tr>
<td></td>
<td>(0.676)</td>
<td>(1.745)</td>
<td>(0.767)</td>
<td>(2.347)</td>
<td>(0.745)</td>
<td>(2.220)</td>
</tr>
<tr>
<td>Constraints 1850 x Frontier Share</td>
<td>-2.728</td>
<td>-2.319</td>
<td>-3.487</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.727)</td>
<td>(3.523)</td>
<td>(3.207)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.177</td>
<td>0.417</td>
<td>0.442</td>
<td>0.091</td>
<td>0.367</td>
<td>0.376</td>
</tr>
<tr>
<td>No. Observations</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

Note: Robust Standard Errors in parenthesis.

All regressions include a constant (omitted).
## Appendix

<table>
<thead>
<tr>
<th>Country</th>
<th>Cartographic Source</th>
<th>Sources for Frontier</th>
<th>Historical References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>Dominion Bureau of Statistics (n.d.)</td>
<td>Silver (1959), Landon (1967)</td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>Butland (1966)</td>
<td>James (1941), Villalobos (1992)</td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>Bureau of Business Research (1975)</td>
<td>Bernstein (1964)</td>
<td></td>
</tr>
<tr>
<td>Venezuela</td>
<td>Butland (1966)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Province/Region</td>
<td>Land Area (km²)</td>
<td>Population</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-----------------</td>
<td>------------</td>
</tr>
<tr>
<td>UGANDA</td>
<td>Central</td>
<td>260,000</td>
<td>4,912,000</td>
</tr>
<tr>
<td></td>
<td>Eastern</td>
<td>380,000</td>
<td>3,690,000</td>
</tr>
<tr>
<td></td>
<td>Western</td>
<td>240,000</td>
<td>2,520,000</td>
</tr>
<tr>
<td></td>
<td>Northern</td>
<td>300,000</td>
<td>1,850,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>Province/Region</th>
<th>Land Area (km²)</th>
<th>Population</th>
<th>Black Population</th>
<th>White Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONGO</td>
<td>Bélier</td>
<td>530,000</td>
<td>1,620,000</td>
<td>640,000</td>
<td>980,000</td>
</tr>
<tr>
<td></td>
<td>Kwango</td>
<td>420,000</td>
<td>2,050,000</td>
<td>820,000</td>
<td>1,230,000</td>
</tr>
<tr>
<td></td>
<td>Luvungi</td>
<td>290,000</td>
<td>1,410,000</td>
<td>560,000</td>
<td>850,000</td>
</tr>
<tr>
<td></td>
<td>shores</td>
<td>360,000</td>
<td>1,890,000</td>
<td>750,000</td>
<td>1,140,000</td>
</tr>
</tbody>
</table>

199
<table>
<thead>
<tr>
<th>Departamento</th>
<th>Población</th>
<th>Municipios</th>
<th>Comuna</th>
<th>Protectorado</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARAGUA</td>
<td>1,431</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CARACAS</td>
<td>2,605</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>COLÓN</td>
<td>4,651</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DOS CAMINOS</td>
<td>5,014</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GUIA</td>
<td>5,353</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HATO CAYO</td>
<td>2,460</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PENSIÓN</td>
<td>4,897</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PICHINCHA</td>
<td>7,112</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TITIRIBI</td>
<td>3,215</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>YARUMBO</td>
<td>5,106</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BODRIK</td>
<td>5,375</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SANTO DOMINGO</td>
<td>7,234</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TURQUESA</td>
<td>5,828</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VITOTOY</td>
<td>6,561</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGUAS CALIENTES</td>
<td>5,020</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BAJA CALIFORNIA</td>
<td>61,457</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CAÑAS</td>
<td>33,400</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAYOLÉ</td>
<td>111,401</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>COWICHÁ</td>
<td>27,604</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUNDUMO</td>
<td>62,600</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DEPENDO</td>
<td>58,897</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>JUTÍPICO</td>
<td>66,130</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KREIZA</td>
<td>59,684</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LAMAS</td>
<td>42,922</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MACAS</td>
<td>79,027</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TUMBES</td>
<td>7,282</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LAMBAYE</td>
<td>56,907</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MARANÉ</td>
<td>101,060</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NAVAL</td>
<td>6,353</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NAUTA</td>
<td>7,262</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RAMÓN</td>
<td>7,188</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BADAYA</td>
<td>3,925</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BARRANQUIA</td>
<td>6,032</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BORICUA</td>
<td>6,628</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BREYES</td>
<td>6,580</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAJAMARCA</td>
<td>6,531</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CANOA</td>
<td>11,100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DISTRITO</td>
<td>9,700</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ORUCA</td>
<td>7,800</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PAREDES</td>
<td>10,700</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PEÑALOY</td>
<td>7,400</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PRIBAGA</td>
<td>13,100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>QUETEJO</td>
<td>7,250</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RUIZ</td>
<td>6,400</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TUMBES</td>
<td>9,200</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NAVAL</td>
<td>7,250</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NAUTA</td>
<td>7,250</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RAMÓN</td>
<td>7,250</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BADAYA</td>
<td>3,925</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BARRANQUIA</td>
<td>6,032</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BORICUA</td>
<td>6,628</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BREYES</td>
<td>6,580</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAJAMARCA</td>
<td>6,531</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CANOA</td>
<td>11,100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DISTRITO</td>
<td>9,700</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ORUCA</td>
<td>7,800</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PAREDES</td>
<td>13,100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PRIBAGA</td>
<td>7,250</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>QUETEJO</td>
<td>7,250</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RUIZ</td>
<td>7,250</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TUMBES</td>
<td>7,250</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>