The Political Economy of Moral Conflict: An Empirical Study of Learning and Law Enforcement under Prohibition*

Camilo García-Jimeno†

This Version: September, 2012

Abstract

The US Prohibition experience shows a remarkable policy reversal. A structural model where law enforcement choices respond to alcohol-related moral views and beliefs about the Prohibition-crime nexus is used to investigate how learning about Prohibition enforcement costs affected the mapping from community preferences to policy outcomes, and was shaped by the outcomes themselves. Estimation is performed through maximum likelihood on Prohibition Era city-level data on police enforcement, crime, and alcohol-related legislation. The model can account for the variation in public opinion changes, and the heterogeneous responses of law enforcement and violence across cities. The paper concludes with counterfactual exercises exploring equilibrium implications of changes in priors, preference polarization, and political environments.

Keywords: Law Enforcement, Learning, Prohibition, Crime

*This paper is 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, Joao Pinho de Mello, Jimena Hurtado, and the participants at MIT's Political Economy Breakfast, the University of Toronto's Applied Micro Seminar, Cornell University's Applied Economics Workshop, and LACEA's Political Economy and Crime and Policy Network Seminars, 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.

†University of Pennsylvania, Department of Economics. gcamilo@sas.upenn.edu.

"Man learns by the disappointment of expectations." Hayek (1960, p. 60)

1 Introduction

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. Its initial introduction was rooted in religious and moral motivations, while its subsequent demise resulted from a widespread dissatisfaction with its effects. This suggests that understanding equilibrium policy change regarding morally motivated issues requires taking into account both individuals´ moral views and their beliefs about the implications of bans on certain activities, behaviors, and expressions. Nevertheless, it is an empirical challenge to distinguish between the effects of changes in moral tastes and changes in beliefs as drivers of policy change because they are likely to be correlated. For example, those who find certain practices abhorrent may also think that banning them can be effective and would have only minor unintended consequences. Moreover, the fact that changes in beliefs about policy effectiveness are endogenous to their outcomes makes the empirical challenge even bigger.

Because moral views are slow-changing, while the uncertainty about policy effects is likely to evolve in response to the policy outcomes themselves, Prohibition's relatively fast backlash and its well known effects on crime make it an ideal setting to study the role of both beliefs and moral views on policy change and policy failure. I 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 public policy, through a feedback between the effects of policies and changing attitudes in response to their effects, modulated by the endogenous extent of their enforcement. In the U.S. Prohibition setting, this mechanism is suggested by John D. Rockefeller, himself a radical prohibitionist, who argued:

> 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 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))

I propose a dynamic structural model that allows me to study the separate roles that moral views and belief changes had on the patterns of Prohibition enforcement and crime, and that explicitly models the endogenous relation between public opinion and crime outcomes. There is heterogeneity in moral views and beliefs, 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 law enforcement choices.

I estimate this model by Conditional Maximum Likelihood, using a dataset of U.S. cities during the period 1911-1936, 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. Then I document how these patterns differed across 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 model explains a large fraction of the variation in police expenditure choices, criminality, and the alcohol-market dynamics. With the model I also estimate the extent to which Prohibition enforcement was 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 four years into Prohibition, the effective alcohol supply fell by around 40%, but rebounded quickly thereafter. Moreover, I estimate that the Prohibition-related homicide rate was increasing in the level of law enforcement.

The structural model also allows for the estimation of several moments of the joint distribution of moral views and prior beliefs. I find that beliefs were optimistic across the distribution of moral views. 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. First, I find that 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. Second, 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. 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.

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.

This paper is related to several research areas. The first studies the determinants of civil liberties. Lagunoff (2001) argues that majorities have incentives in the present to weaken the enforcement technologies available, which could otherwise be used against themselves in the future. Haider-Markel and Meier (1996) emphasize that the polarizing nature of morally-charged issues makes them prone to interest-group politics. 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. This paper highlights that when legal restrictions on individual liberties have potentially uncertain side effects, these effects can become a source of opinion and policy change.

The literature on "crime and punishment" pioneered by Becker (1968) has focused on understanding the determinants of crime enforcement and the effects of law enforcement on the equilibrium levels of illegal activity by emphasizing that *de jure* prohibitions require concomitant *de facto* enforcement. This paper highlights that punishment is a social choice. As such, what society defines as crime is endogenous to its willingness to enforce its own legal standards. 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 policy 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. Buera et al. (2011) 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. In the same spirit, Mukand and Rodrik (2005) 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, where incentives for experimentation are limited by the trade-off between learning about the effects of policies and remaining pivotal.

Finally, this paper also contributes to the literature on the determinants of crime (See Dills et al. (2008) for a recent survey). The literature has stressed demographics, criminal policy, access to firearms, investment in policing, inequality, and the economic cycle (Levitt (2004); Donohue and Levitt (2001); Dills et al. (2008)). Miron (1999) 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. A recent contribution is Owens (2011), who directly looks at the introduction of dry laws on state-level crime during the Prohibition period. 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. 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

Nation-wide alcohol Prohibition was written into the US 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[1]. 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, during 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 did not constitute a majority. Their political success was due to their pivotal character in the competitive context of bipartisan politics, based on strong local campaigning and national lobbying, and on the intensive use of referenda initiatives. Republicans and Democrats were frequently so evenly divided that a switch of the temperance vote could decide local elections. Prohibitionist groups were able to become pivotal even in the within party races of the Democratic-dominated South. [2]

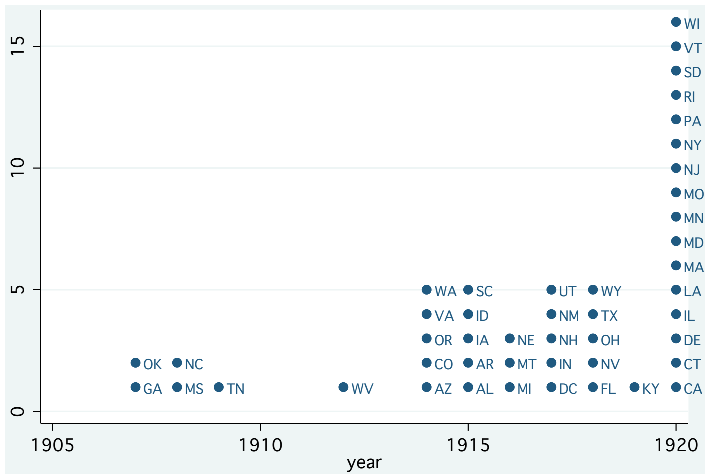
Key to the political success of the drys was their lack partisan aligniment. There was disagreement on the issue within Southern Democrats too, as a faction of the party believed that allowing the Federal government to make decisions regarding Prohibition could be a step to further undermine Southern 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 element to explain the dry

[1] 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.

[2] 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).

success was its gradual approach. Local option measures were followed by state-wide legislation. Right before the 18th Amendment was adopted, almost 80% of U.S. counties were already under some form of Prohibition. Figure 1 shows the dates of state-adoption of Prohibition legislation. 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 [3].

Figure 1: Timing of State Adoption of Prohibition



Before Constitutonal Prohibition, enforcement of alcohol laws in states under Prohibition was usually mild. In dry communities it was redundant, while in wet communities it was relatively ignored. Most alcohol consumption took place in saloons and other public spaces, which made public intoxication a widespread phenomenon (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)).

Although the passage of the 18th Amendment and its enforcement law (the Volstead Act)[4] 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. Congress, expecting local and state cooperation 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 [5].

[3] 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.

[4] President Wilson vetoed the Volstead Act, and his veto was overridden by Congress.

[5] 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)).

Data on Federal enforcement outcomes during the 1920s shows similar trends across the four main U.S. regions. 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, suggesting a sharp fall in the enforcement activities against producers. Indeed, by 1928 several states had already repealed their own enforcement legislation. The most prominent case was that of New York, which in 1923, repealed the state enforcement law. Alfred Smith, the Democratic Presidential candidate in the 1928 election, was then the Governor of New York. 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)).

To have an idea of the limited extent of law enforcement at the federal level, in 1929-1930, total liquor seizures in the U.S., including spirits, malts, wines, cider, mash, and pomace, were approximately 74 million gallons (U.S. Bureauof Prohibition (Several Years), and Wickersham-Commission (1928-1931)). 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, federal enforcement looks insignificant.

Most law enforcement relied on local efforts, not only because of the inherent difficulties in enforcing alcohol restrictions throughout the country, which limited the federal law enforcement stategies 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 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 drys saw the problem in the ineptitude and corruption of enforcers, wet 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 millon 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)).

The extent of Prohibition enforcement 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),

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)).

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 (Szymansky (2003, p. 184), Kyvig (1979, p. 25), Tydings (1930, p. 127)). Judicial leniency was even institutionalized through the so-called "bargain days", which arose 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 prosecutions[6].

The effects of Prohibition also varied across 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 areas[7]. 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 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. The public opinion shift in other regions took place at a slower pace, and more in response to the observable increase in criminality. Initially dry individuals, who were morally compeled 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. Neither was the fall in support for the policy. The Democratic party, out of power throughout the 1920s, managed to capture most of the rise in anti-Prohibitionist sentiment, partly explaining Franklin D. Roosevelt's victory in the 1932 Presidential election. 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 advertising campaigns in 1928, 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[8]. Among the rest of the states, some instituted systems of "state operation", in which the state directly controlled the distribution of alcohol; others just imposed some regulation over a free market (Harrison and Laine (1936, p. 43)).

[6] I collected the data on judicial prosecutions at the judicial district level for the period 1915-1933 directly from the Attorney General Annual Reports.

[7] "New 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 (1928-1931, Box 13-2, Prohibition Survey of New York, p.2)).

[8] These 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))

3 Data and Summary Statistics

I used several sources to collect city-level homicide-rate data (up to 93 cities) and drunkenness arrests data (up to 573 cities) covering the period 1911-1936. Due to my focus on local law enforcement, I also obtained 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 250 cities covering 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[9].

City and county-level data on demographic characteristics are taken from the decennial population censuses. I focus on the age distribution, the ethnicity distribution[10], and total population, from the 1910-1940 Censuses. I also 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. 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).

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 were usually ballot measures proposed to the citizens to approve or repeal liquor laws, or ammend the state constitutions. 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[11]. 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.

Finally, alcohol-related legislation across states comes from several historical sources based on which 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. (See Online Appendix 5 for a detailed description of the sources).

Table 1 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. It 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.

[9] 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.

[10] I focus on the distribution of the population between native white, foreign white, and black individuals.

[11] The main caveat here is that 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).

Table 1: Summary Statistics

Summary Statistics	Region*							
	Midwest		Northeast		South		West	
Dry Religions								
% Baptist (1916)	12.066		8.601		29.660		16.149	
	(8.559)		(5.901)		(13.921)		(9.245)	
% Evangelical (1916)	2.623		0.647		0.986		0.852	
	(2.083)		(1.393)		(1.496)		(0.510)	
% Mormon (1916)	0.298		0.057		0.160		5.198	
	(0.578)		(0.140)		(0.413)		(18.856)	
% Methodist/Episcopal (1916)	11.666		10.269		24.825		13.740	
	(6.356)		(4.740)		(9.855)		(6.469)	
% Presbyterian (1916)	4.779		4.471		6.060		8.068	
	(2.642)		(3.521)		(3.246)		(4.743)	
Wet Religions								
% Eastern Orthodox (1916)	1.139		1.272		0.437		1.890	
	(1.443)		(1.461)		(0.591)		(2.222)	
% Jewish (1916)	1.741		2.771		1.716		1.887	
	(1.130)		(2.515)		(1.303)		(1.033)	
% Lutheran (1916)	7.677		3.510		2.521		3.181	
	(5.708)		(3.678)		(2.119)		(1.944)	
% Catholic (1916)	55.798		66.631		31.496		47.494	
	(16.245)		(12.568)		(21.798)		(20.723)	
Demographics								
% Black (1910)	3.104		2.325		26.525		1.336	
	(2.823)		(2.057)		(12.398)		(0.951)	
% Foreign White (1910)	24.573		32.618		7.928		23.083	
	(10.557)		(8.664)		(5.705)		(5.364)	
% Native White (1910)	72.220		64.945		65.412		73.210	
	(9.036)		(8.170)		(9.816)		(5.610)	
% Ages 15-24 (1910)	20.840		20.041		21.050		18.970	
	(0.909)		(1.010)		(1.166)		(0.909)	
% Ages 25-44 (1910)	34.258		33.515		34.159		38.519	
	(1.610)		(1.204)		(2.163)		(3.087)	
Legislation								
Number of Dry Laws (1919)	6.709		3.376		5.999		9.303	
	(3.136)		(2.014)		(1.696)		(3.310)	
Number of Years Under Prohibition**	15.925		15.018		18.035		15.844	
	(4.443)		(0.169)		(4.080)		(1.253)	
Outcomes	1910s	1920s	1910s	1920s	1910s	1920s	1910s	1920s
Per Capita Police Expenditure (1913 prices)	1.541	1.793	1.977	2.312	1.376	1.716	1.511	1.796
	(0.726)	(0.759)	(0.732)	(0.798)	(0.513)	(0.800)	(0.687)	(0.699)
Police Expenditure Share	0.108	0.092	0.112	0.096	0.123	0.112	0.087	0.081
	(0.032)	(0.026)	(0.021)	(0.020)	(0.024)	(0.026)	(0.026)	(0.017)
Drunkenness Arrest Rate (per 1,000)	16.000	14.459	16.653	12.132	18.273	18.606	22.560	13.963
	(11.191)	(8.591)	(18.875)	(13.081)	(11.525)	(10.786)	(14.801)	(6.241)
Homicide Rate (per 100,000)	10.807	18.124	5.368	10.076	22.849	28.132	9.897	11.620
	(5.730)	(7.560)	(1.776)	(3.633)	(18.085)	(17.150)	(3.375)	(3.379)
% Anti-Prohibition vote share***	0.518	0.723	0.523	0.826	0.467	0.577	0.457	0.678
	(0.169)	(0.168)	(0.103)	(0.129)	(0.166)	(0.220)	(0.138)	(0.116)

*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

For the religious distribution, I present summary statistics from the 1916 Census of Religions. 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 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, Western cities present the most uniform religious membership distribution. In contrast, religious membership in Southern cities was heavily dry, while in the Midwest and Northeast wet religions were majoritarian. The Table shows that while 26% of the population in Southern cities was black, this ethnic group represented only between 1.3 and 3.1 percent in all other regions. 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 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. 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 some differences between regions. In Southern cities, average arrests were very similar in the 1910s and 1920s. In contrast, cities in the West 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.

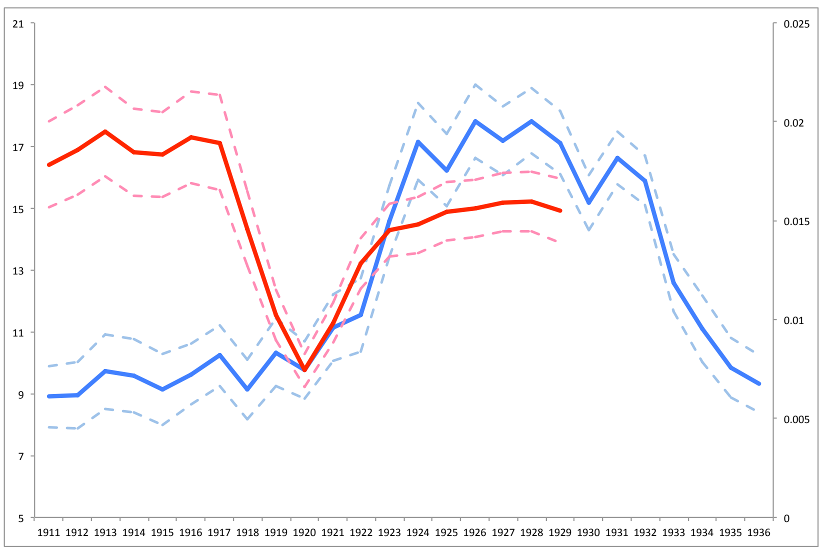
4 Some Reduced Form Results

4.1 Crime, Law Enforcement, and the Duration of Prohibition

I begin the empirical analysis by presenting two main features in the data, suggestive of a dynamic feedback between crime outcomes and law enforcement choices operating through a channel of learning about the effects of Prohibition enforcement. First I show that the co-evolution of crime and law enforcement over time suggests initial optimism and subsequent pessimism about the social costs of enforcing the policy. Then I present the basic cross-sectional patterns of public opinion change. Despite being initially less favorable to the policy, cities with larger wet constituencies faced both larger crime increases and larger public opinion shifts against it. This patter is also suggestive of the importance of learning as a source of policy preference change.

I first focus on the timing of Prohibition adoption across states, and on three outcome variables: The homicide rate as a measure of criminality, the drunkenness arrest rate as a way to look at the alcohol market dynamics, and police expenditure as a measure of law enforcement. Figure 2 shows that the advent of Prohibition saw a sharp increase in crime, as measured by the homicide rate (blue line). Nevertheless, the increase 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.

Figure 2: Homicide Rate and Drunkenness Arrest Rate in U.S. Cities, 1911-1936



Blue: Homicide Rate (per 100,000). Red: Drunkenness Arrest Rate (per 1,000). Dashed lines represent 95% confidence intervals.

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 also 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 40% 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, occured staggeredly across states during the first two decades of the century.

Time under Prohibition is a convenient reduced-form way to examine its time-varying effects for several reasons. First, because of the alcohol supply dynamics. After Prohibition was adopted, the legal market for alcohol was closed on impact, leading to a large shock on the availability of liquor. The black market required time to develop smuggling networks, hidden production facilities, and criminal organizations supporting it. Finally, law enforcement was a key channel through which Prohibition had an impact on the development of criminality, which responded to communities' beliefs about the policy. The evolution of these beliefs over time was also a dynamic force shaping the time-varying effects of Prohibition. To take a first look at short-run and long-run effects of Prohibition, I start by estimating fixed-effects models of the form

$$y_{ct} = \alpha_c + \beta_t + \sum_{\tau=1}^{k} \delta_\tau D_{c\tau} + \boldsymbol{\gamma}' \boldsymbol{X}_{ct} + \varepsilon_{ct} \tag{1}$$

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 α_c are city-specific effects, the β_t are year-effects, and the $D_{c\tau}$ are indicator variables for each cumulative number of years under Prohibition[12]. The vector $\boldsymbol{X}_{ct}$ 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 δ_τ, the time-varying effects of Prohibition. Since this model looks only at within-city variation over time, δ_τ 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 τ years. Standard errors reported are robust to arbitrary heteroskedasticity and clustered at the city level.

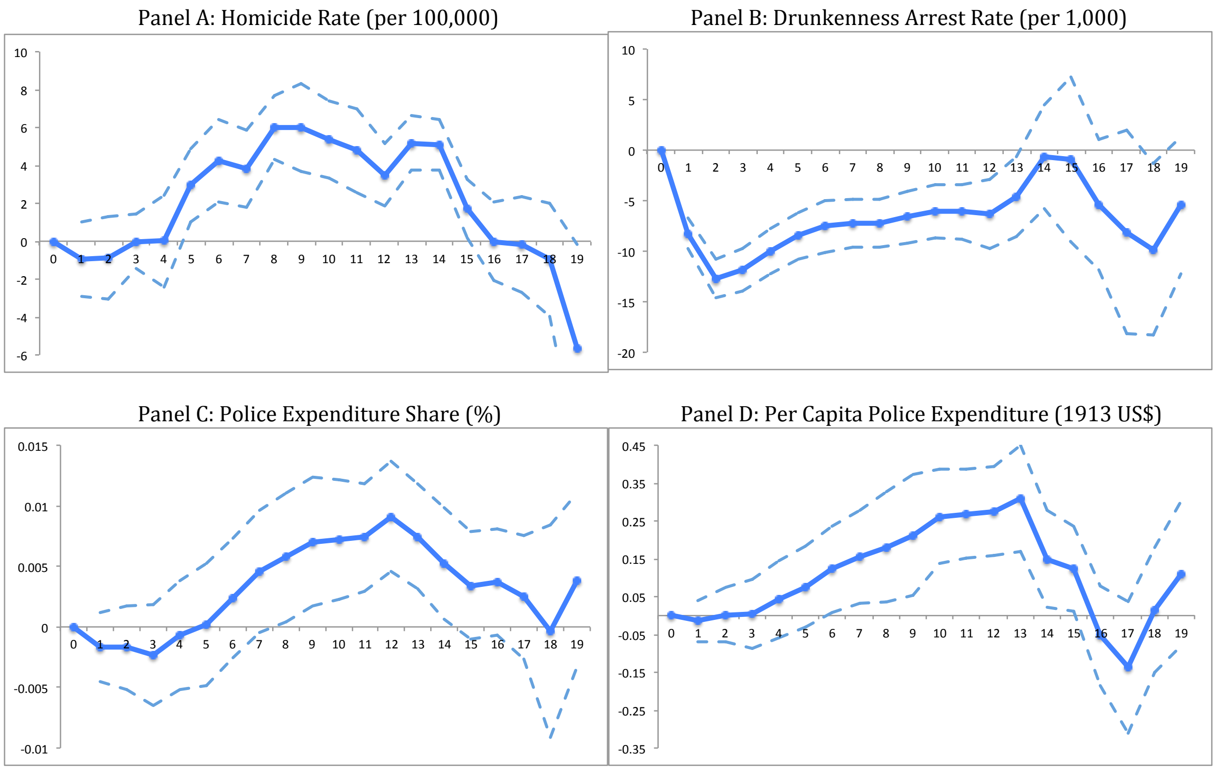
For brevity and ease of illustration, here I present results for an unbalanced panel covering the 1911-1936 period, excluding cities for which there are less than ten years of data for drunkenness arrests or police expenditure, or less than eight years of homicide rate data (See Online Appendix 4 for results using alternative samples). Figure 3 graphs the estimated δ_τ's of the baseline specification with no year effects for the different outcome variables. The homicide rate (Panel A) 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. Then it starts slowly to fall back to a level similar to the pre-Prohibition average. The set of cities experiencing lengthier periods under Prohibition shrinks over time, so late δ_τ's are less precisely estimated. At its peak, cities were on average experiencing 4 homicides per $100,000$ more than before Prohibition was introduced (s.e.$=1.1$).

Analogous regression results for drunkenness arrests (Panel B) provide a complementary picture. The figure illustrates the sharp fall in drunkenness arrests during the first two years after a city was under Prohibition, consistent with the impact closing of most of the supply sources of alcohol which, during this period, were to a large extent domestic. During the third year under Prohibition, drunkenness arrests attain a minimum. The estimated coefficient for δ_3 is -11.83 (s.e.$=1.1$), implying a 55% ($=11.83/19$) contraction of the alcohol supply in the absence of changes in law enforcement or demand. The figure also illustrates the steady recovery of the alcohol market, (assuming arrest

[12] In the sample τ 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.

intensity did not change significantly throughout Prohibition). Approximately fifteen years into Prohibition, drunkenness arrests are indistinguishable from Pre-Prohibition levels [13].

Figure 3: δ_τ's from equation (1)



δ_τ coefficient estimates from regression (1). Dashed lines represent 95% confidence intervals for the coefficient estimates.

These patterns 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 argued 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, panels C and D in Figure 3 present the estimates of the δ_τ's both for the police share and for the 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 after a few years into 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

[13]The identification assumption here is that the introduction of Prohibition did not also induce changes in individuals' preferences over alcohol consumption. As an effort to check how reasonable this assumption is, Online Appendix 4 presents some evidence exploiting variation in the availability of neighboring alcohol markets. The evidence there is consistent with little change in demand after the introduction of Prohibition.

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[14].

Although the Prohibitionist legal standard was imposed to cities at the state level, Online Appendix 4 presents some robustness results exploring potential correlates of the timing of adoption of Prohibition, such as the availability of neighboring alcohol supply sources, pre-Prohibition state-level alcohol-related legislation, and women's suffrage legislation. The evidence does not suggest that alternative legislation was driving the patterns described above.

4.2 Moral heterogeneity and Public Opinion

The result of the trends described above was a massive public opinion backlash, best illustrated by Figure 4A. It presents the distribution of wet vote shares in alcohol-related referenda, available at the county level for most of the US states, taking place in different years during the 1910s-1930s[15]. I focused on finding for each state, electoral returns on a liquor referendum taking place prior to the introduction of Prohibition in the State (pre-Prohibition period), and for a year in the later Prohibition period or after the repeal of federal Prohibition (post-Prohibition period). In the pre-Prohibition referenda, the 75th percentile of the distribution of wet vote shares is 0.5. Thus, Prohibition had majoritarian support in three quarters of the counties. In the post-Prohibition period, only 35% of counties had majorities favoring Prohibition. A comparison of both histograms also suggests a spreadout in the distribution of public opinion.

To explore this public opinion shift and spreadout in more depth, I constructed a proxy for the "wet share" in the population, μ_{ct}, as the sum of the fractions of the population in any of the "wet" religions, the share of non-native white individuals, and the share of the 15-44 years old population[16]. There is fairly widespread consensus that Baptist, Evangelical, Methodist, Mormon, and Presbyterian religious ascriptions were more favorable to Prohibition, while Catholic, Orthodox, Jewish, and Lutheran communities were more sympathetic to 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); Szymansky (2003); Blocker (1989); Asbury (1950)). I define μ as:

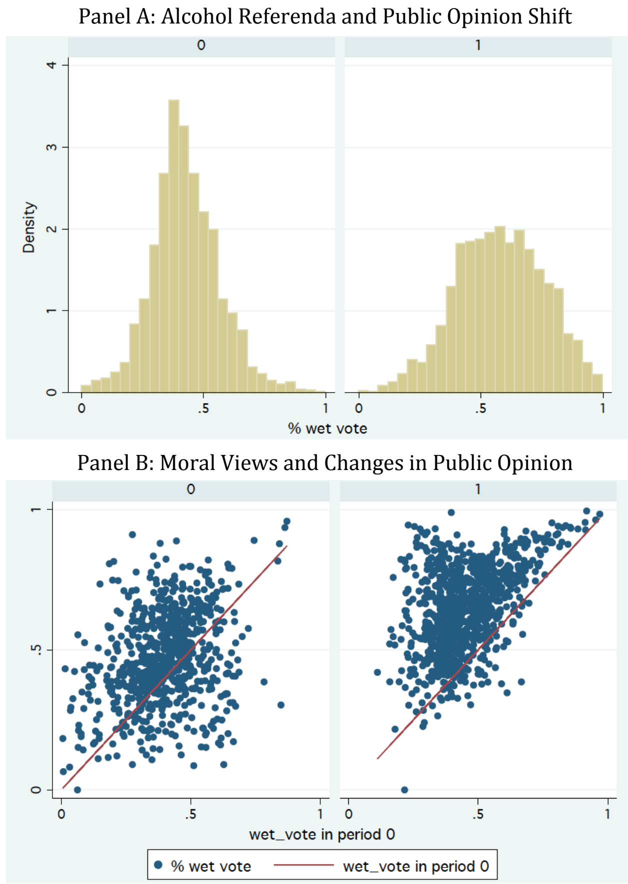
$$\begin{aligned}\mu_{ct} = \tfrac{1}{3}(1 - \%Baptist_{ct} - \%Evangelical_{ct} - \%Methodist_{ct} - \%Mormon_{ct} - \%Presbyterian_{ct}) \\ + \tfrac{1}{3}(1 - \%NativeWhite_{ct}) + \tfrac{1}{3}(\%PopulationAges15-44_{ct})\end{aligned} \tag{2}$$

[14] 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. All states except Maryland adopted state-level enforcement legislation right after the passage of the Volstead Act, complying with the shared-enforcement responsibilities established by the 18th Amendment. Throughout the 1920s several states repealed their state-enforcement laws. The state of New York took the lead in 1923. It was followed by Montana in 1925, Nevada and Wisconsin in 1928, Massachussetts in 1930, and Arizona, California, Colorado, Louisiana, Michigan, North Dakota, New Jersey, Oregon, and Washington in 1931.

[15] Data is available at the city level for Massachusetts and Connecticut.

[16] I normalize this variable dividing by 3, the total measure of the religious, ethnicity, and age distributions.

Figure 4



County-level data. The wet vote share is computed as the fraction of the total vote favoring the wet side of the referendum under consideration. Pre-prohibition data are taken from the latest state-level liquor referendum taking place prior to the adoption of Prohibition in the state. Post-Prohibition data are taken from the 21st Amendment Constitutional Convention election results.

A break-up of the sample between counties below and above the median value of μ in 1911 (0.355)[17] reveals a suggestive pattern illustrated in Figure 4B. It plots the pre-Prohibition and post-prohibition wet vote shares in the horizontal and vertical axes respectively, together with a 45 degree line, with counties below the median in the left panel, and counties above the median in the right panel. The largest public opinion shifts took place in communities above median μ. In contrast, pre-Prohibition vote shares remain a strong predictor of post-prohibition vote shares for below-median counties.

The differentially larger shift against Prohibition in initially more anti-Prohibitionist communities could be explained by selection bias because the most anti-Prohibitionist states never held pre-Prohibition referenda. It could also be the result of time-varying state-level shocks correlated with demographics, or to specific features of the referenda taking place in different states. Changes in the demographic and religious composition of these communities themselves could also explain the differential public opinion shift. Alternatively, it could be the result of differential changes in beliefs about the success of the policy. Large differences in priors could lead to differences in learning

[17] I computed μ for each county directly from equation 2 using the 1916 and 1926 Census of Religions and the 1920 and 1930 Population Censuses (county-level age distribution from the 1910 census is unavailable).

speed. On ther other hand, these trends could be observed even under similar learning speeds across communities if these varied in their Prohibition-enforcement costs, or if preferences over public policy were very inelastic to learning due to very extreme moral views by the driest of communities.

Disentanging these channels is not possible in a reduced-form framework. Nevertheless, Online Appendix 4 presents some results suggesting that neither selection, demographic trends nor referenda features were important. It shows that initial differences in the demographic/preference profile of communities was a key determinant of the different experiences under Prohibition. The Appendix also presents additional reduced form evidence showing that the largest increases in the homicide rate took place precisely in the cities with larger wet populations. This suggests that part of their differentially larger public opinion shift against Prohibition was due to worse outcomes under the policy. The structural model will allow me to examine the separate roles of moral tastes, learning, and their interaction.

Two main facts emerge from the reduced-form results. First, a tandem rise-and-fall pattern in crime and law enforcement across cities. Second, a differentially larger shift in public opinion against Prohibition coupled with larger crime increases in initially less Prohibitionist cities. Both are consistent with changes in policy preferences responding to learning about the costs of Prohibition enforcement. They suggest an experimentation period, taken back after its costs and ineffectiveness became evident.

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

In this section I develop a political economy learning model of Prohibition enforcement which provides enough structure to be directly estimated. Society is made up of a many communities $c = 1, 2, ...$, in discrete time. Community c is populated by a continuum measure 1 of adult citizens indexed by i. Each period $t = 0, 1, 2, ...$, 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[18].

In addition, society as a whole 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. 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. 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

[18] Throughout this section I drop the community indices c, since no confusion arises. In section 6 I specify which parameters are city-specific for estimation purposes.

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):

$$f_\mu(\mu; a, b) = \frac{\mu^{a-1}(1-\mu)^{b-1}}{\int v^{a-1}(1-v)^{b-1}dv}, \quad a, b > 0 \tag{3}$$

Individuals know the parameters of the distribution, but do not observe the draw directly. 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 assume it is inherited from parent to child. Individuals have common values about consumption of a public good G, and crime, but there is heterogeneity in prior beliefs about how the introduction of Prohibition might impact crime. Thus, conflicting views over Prohibition arise not only from differences in individual moral stands (tastes), but also from informational differences. Nevertheless, these are correlated in the population 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 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 ζ^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 ξ^i is an individual-specific bias. (ζ^i, ξ^i) is drawn from a joint-normal distribution

$$\begin{pmatrix} \zeta^i \\ \xi^i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\zeta^2 & \rho\sigma_\zeta\sigma_\xi \\ \rho\sigma_\zeta\sigma_\xi & \sigma_\xi^2 \end{pmatrix}\right) \tag{4}$$

While moral views are not subject to updating with the arrival of new information, beliefs evolve through rational updating as individuals receive new information. In the context of Prohibition, it is natural to think of crime as the source of information about θ. 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 (ζ^i, ξ^i) drawn from the same distribution.

The expected utility of a citizen is given by

$$E_t U^i(h_t^i, A_t, G_t, q_t | P_t) = E\left[1_{\{i \in W_t\}} h_t^i - z^i A_t + V(G_t) - q_t\right] \tag{5}$$

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.

5.1 The Alcohol Market

The price of consuming a unit of liquor is normalized to zero, but individuals must engage in a search. The probability of successful search is a decreasing function of the level of Prohibition enforcement chosen by the community[19]. 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.

The introduction of Prohibition makes legal alcohol unavailable making search costlier. I allow the probability of a successful search to also depend on the amount of time the community has been under Prohibition, τ_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 $Pr(h_t^i = 1|P_t = 1) = k(\tau_t)exp(-e_t)$, where

$$k(\tau_t) = 1 - \lambda\tau_t exp(-\kappa\tau_t) \tag{6}$$

with $\kappa, \lambda > 0$[20]. It follows that aggregate alcohol consumption is

$$A_t(e_t) = \int_{i \in W_t} k(\tau_t)exp(-e_t)di = \mu_t k(\tau_t)exp(-e_t) \tag{7}$$

so that during the τth 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 prefer 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.

5.2 Crime, Prohibition, and Law Enforcement

I allow crime to be related to alcohol consumption at any point in time by assuming that baseline crime is proportional to the alcohol market size. This gives room in the model for Prohibitionists' claims about the ill effects of alcohol consumption. I also allow it to vary with the size of the alcohol market. 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. Since Prohibition opened the door for a new source of

[19]Recall that under no Prohinition dry laws were in place. These restricted the availability of liquor by regulating the alcohol market along different dimensions.

[20]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 (λ), and the speed at which the alcohol market bounces back (κ), 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 8 for $\kappa = 0.23$ and $\lambda = 0.3$ (the MLE estimates). This curve has its unique minimum at $\tau_t = 1/\kappa$, so that κ is also the inverse of the time at which the alcohol market reaches its mimimun size. I also impose the condition $\kappa exp(1) > \lambda$, which is necessary and sufficient for $k(\tau_t)$ to be everywhere positve. A comparison of Figures 2b and 8 illustrates why the functional form choice in 6 is likely to be appropriate.

criminality related to the alcohol market, I allow for a Prohibition-specific relationship between crime and alcohol markets. Formally,

$$\begin{aligned} q_t &= \Theta_S + A(e_t) + P_t\theta[A_t(e_t = 0) - A_t(e_t)] + \varepsilon_t \\ &= \Theta_S + k(\tau_t)\mu_t exp(-e_t) + P_t\theta k(\tau_t)\mu_t[1 - exp(-e_t)] + \varepsilon_t \end{aligned} \tag{8}$$

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 Θ_S. $\varepsilon_t \sim N(0, \sigma_q^2)$ is an iid normally distributed shock. Equation (8) captures the two main channels from the alcohol markt 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 [21]. θ is 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 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 the Prohibition-related component of crime is zero if $e_t = 0$, or under no Prohibition. Also, as $e_t \to \infty$, Prohibition-related crime $\to \theta_c k(\tau_t)\mu_t$. 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 θ.

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 $\chi > 0$[22]. The drunkenness arrest rate is thus:

$$d_t = Pr(Arrest|e_t)A_t(e_t) = \frac{\mu_t k(\tau_t)}{\chi + exp(e_t)} \tag{9}$$

This equation holds both under no Prohibition and under Prohibition. The equilibrium drunkenness arrest rate is a decreasing function of law enforcement. Equation (9) highlights that variation in the drunkenness arrest rate can come from changes in the size of the alcohol market, (the wet share μ_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.

[21] In a classic Sociology paper, Paul Goldstein discusses the different channels from drug use to violence. He identifies two sources of criminality in a no Prohibition environment: psychopharmacological and economycally compulsive: The former is due to violent behaviors induced by alcohol consumption. The latter is due to the use of violent inherent in illegal makets." (Goldstein (1985, pp.146-149))

[22] 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 χ.

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 α_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,

$$G_t = 1 - p_t \tag{10}$$

5.3 Learning and the Timing of Events

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 θ according to Bayes' rule. This occurs as follows. First, each child learns her parent's belief θ^i_{t-1}. In the first year under Prohibition ($\tau_{t-1} = 1$), child i knows that $\theta^i_0 = B + \xi^i$ (of course, she does not observe B or ξ^i separately), and knows that $\xi^i \sim N(0, \sigma^2_\xi)$ is the marginal distribution of biases in the population. As a result, child i's prior about θ is given by $\theta^i_{t-1} \sim N(\theta^i_0, \sigma^2_\xi)$.

From equation (9), after the child has observed d_{t-1} and p_{t-1}, she can perfectly back-up the realization of μ_{t-1}. Thus, in the public signal $q_{t-1} = \Theta_S + \mu_{t-1}k(\tau_{t-1})exp(-\alpha_{t-1}p_{t-1}) + \theta k(\tau_{t-1})\mu_{t-1}[1 - exp(-\alpha_{t-1}p_{t-1})] + \varepsilon_{t-1}$, the only remaining uncertainty comprises the true value of θ and the distribution of ε_{t-1}. It follows that Bayesian individuals' posteriors about θ 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 distributed $N\left(\overline{\theta}^i_t, \Omega_t\right)$, where

$$\overline{\theta}^i_t \equiv \Omega_t \frac{1}{\sigma^2_\xi}\theta^i_0 + \Omega_t \frac{1}{\sigma^2_q}\sum_{s=s_0}^{t-1}[q_s - \Theta_S - \mu_s k(\tau_s)exp(-\alpha_s p_s)]\omega_s = \Omega_t \frac{1}{\sigma^2_\xi}\xi^i + \Omega_t \overline{\theta}^C_t \tag{11}$$

is the posterior mean, and $\Omega_t \equiv \frac{1}{\frac{1}{\sigma^2_\xi} + \frac{1}{\sigma^2_q}\sum_{s=s_0}^{t-1}\omega^2_s}$ is the posterior variance, and where I express the posterior mean more compactly by defining $\overline{\theta}^C_t \equiv \frac{1}{\sigma^2_\xi}B + \frac{1}{\sigma^2_q}\sum_{s=s_0}^{t-1}[q_s - \Theta_S - \mu_s k(\tau_s)exp(-\alpha_s p_s)]\omega_s$ to be the common component of beliefs (shared by all individuals in the community). s_0 is the first year in which community c is under Prohibition, and $\omega_t \equiv k(\tau_t)\mu_t[1 - exp(-\alpha_t p_t)]$ is a measure of the degree of informativeness of the signal [23]. This posterior will be the relevant measure with

[23]Equation (11) above highlights the convenience of assuming a normal learning structure, which results in posterior conditional expectations being linear in the signal sequence, and making estimation relatively straightforward. Although this seems to be a very restrictive set of assumptions about the information structure and the cognitive requirements of individuals, 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

respect to which individual i will evaluate her expected utility under different law enforcement policy alternatives. The posterior mean belief at any time t is a weighted average of the prior mean and the history of crime realizations, where weights depend on their relative precisions and the informativeness of each signal. The degree of informativeness depends, in turn, on the extent of law enforcement originating the signal. Equation (11) shows that individual belief sequences can be analytically decomposed into a common component, shared by all individuals in the community 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 is convenient. When the precision of the distribution of prior biases is low (as measured by $1/\sigma_\xi^2$), 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 $(1/\sigma_q^2)$ induces a Bayesian individual to put more weight on her prior. Moreover, since individuals know the DGP up to the uncertainty about θ, they optimally use the information on law enforcement to decide how much weight to give to the crime signal.

The stochastic process in (11) is a bounded martingale, and as such, the $\{\theta_t^i\}$ converge almost surely as $t \to \infty$. Moreover, because the true distribution is absolutely continuous with respect to the prior, the process will converge to the true θ for any infinite sequence of positive $\{\mu_t, e_t\}_{t=s_0}^n$. The speed of convergence 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 learn.

5.4 Political Equilibrium and Preferences over Law Enforcement

Taking a look at the problem by replacing the successful-search probability and equations (7), (8), and (10) into (5), indirect preferences under Prohibition are explicitly obtained. From the first order condition, the preferred police enforcement of individual i under Prohibition is given by equation (12) below, which makes use of equation (11) (See Appendix 1).

$$p_t^*(\zeta^i, \xi^i) = \frac{1}{\alpha_t - 1}\left\{ ln\left[\alpha_t k(\tau_t)\right] + ln\left[\frac{a}{a+b}\left(z - \Omega_t \overline{\theta}_t^C P_t + 1\right) + \frac{a}{a+b}(\zeta^i - \Omega_t \frac{1}{\sigma_\xi^2}\xi^i P_t) - 1_{\{i \in W_t\}}\right] - 1\right\} \tag{12}$$

If the expression inside $ln[]$ is negative, $p_t^*(\zeta^i, \xi^i) = 0$. This expression follows from the fact that μ_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 θ 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

signal sequence $\{q_i\}_{s_0=0}^{t-1}$, their prediction of the conditional mean will exactly match the posterior mean under normal updating, no matter the true data generating process (See Vives (2010, p. 379)).

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 (12) 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 α_t though, ideal policing is falling in α_t.

During Prohibition times, an individual's mean belief about θ also matters. Individuals must now include the increased criminality induced by Prohibition enforcement in their optimal trade-off regarding police expenditure. Equation (12) 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 ζ^i, but also of their dynasty-specific belief bias ξ^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' when all adult citizens vote sincerely given their current beliefs $F_t^i(\theta)$, sequences of homicide and drunkenness arrest rates $\{q_t\}_{t=0}^{\infty}$, $\{d_t\}_{t=0}^{\infty}$ given by (8) and (9), and a sequence of belief distributions $\{F_t^i(\theta)\}_{t=0}^{\infty}$ for each dynasty i, which are updated every period according to Bayes' rule.*

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 $F_t^i(\theta)$ $\forall i \in [0,1]$, and a legal standard vector (α_t, τ_t, P_t), there is a unique equilibrium level of law enforcement p_t given by*

$$p_t^* = \frac{1}{\alpha_t - 1}\left\{ ln\left[\alpha_t k(\tau_t)\right] + ln\left[\frac{a}{a+b}\left(z - P_t \Omega_t \overline{\theta}_t^C + 1\right) + (1-P_t)\varrho_N^{med} + P_t \varrho_P^{med}\right] - 1\right\} \tag{13}$$

where ϱ_N^{med} and ϱ_P^{med} are random variables whose densities $f_{\varrho_N^{med}}(\varrho_N^{med}; a, b, \sigma_{\varrho N})$ and $f_{\varrho_P^{med}}(\varrho_P^{med}; a_c, b, \sigma_{\varrho P t})$ are continuous and positive over the interval $[-1, 0]$.

Proof. See Appendix 2. □

As the proof of Proposition 1 shows, ϱ_N^i and ϱ_P^i 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 μ_t. In Appendix 2 I show that the equilibrium level of law enforcement is determined by the median voter's value of ϱ_j^i. Because the wet share is itself a beta-distributed random variable, ϱ_j^{med} is also a random variable whose equilibrium density $f_{\varrho_j^{med}}(\varrho_{Nj}^{med}; a, b, \sigma_{\varrho_j})$ is continuous and takes positive values over the interval $[-1, 0]$. As $\mu_t \to 1$, $\varrho_j^{med} \to -1$, and as $\mu_t \to 0$, $\varrho_j^{med} \to 0$. 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[24].

5.5 Predictions

Equation (13) makes both time series and cross sectional predictions about the equilibrium dynamics of law enforcement. Over time, law enforcement should vary with the size of the alcohol supply $k(\tau_t)$, for two reasons. First, because the equilibrium size of the alcohol market depends on supply and law enforcement. As supply contracts (expands), less (more) law enforcement is required to maintain a given alcohol market size. Second, because Prohibition-related crime also depends on the alcohol market size and law enforcement. If beliefs are that $\theta > 0$, as supply contracts (expands), less (more) law enforcement is required to maintain a given level of crime. Law enforcement also responds to the difference between average moral tastes and mean beliefs $(z - \Omega_t \overline{\theta}_t^C)$ over time. For communities starting with optimistic beliefs $(\Omega_t \overline{\theta}_t^C < \theta)$, learning over time will reduce this difference making law enforcement less attractive as its expected cost (crime) increases relative to its benefit (a smaller alcohol market and thus, a reduced moral externality). Moreover, the speed of learning depends on ω_t so that the size of the alcohol supply and the extent of law enforcement will also affect how fast $(z - \Omega_t \overline{\theta}_t^C)$ changes. For example, periods of low alcohol supply should lead to slower learning because crime signals are not very informative. Finally, the change in Prohibition status predicts a difference in the variance of law enforcement choices due to the change in the distribution from which median unobserved heterogeneity ϱ^{med} is drawn. If ρ, the correlation between idiosyncratic moral tastes and prior beliefs is large in magnitude, draws of ϱ_P^{med} will be on average more extreme than draws of ϱ_N^{med}, leading to less variation in law enforcement during Prohibition.

Two key cross sectional predictions are suggested by equation (13). Willingness to enforce should vary across cities with different sizes of their wet population $(a/(a+b))$. Nevertheless, because wet population size and beliefs interact, the sign of the difference should change as communities undergo learning. On the other hand, cities with larger religiously motivated constituencies (larger z) should also be willing to enforce more. Hence, although the prediction about the relationship between demographics and law enforcement is ambiguous early on during Prohibition, once significant amount of learning has taken place, demographically more Prohibitionist cities should be willing to enforce more. Finally, changes in beliefs should lead to smaller changes in law enforcement in cities with larger religiously motivated constituencies.

[24]For a discussion of the main assumptions behind the modeling choices, see Online Appendix 6

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):

$$q_{ct} = \Theta_S + k(\tau_{ct})\mu_{ct}exp(-\alpha_{ct}p_{ct}) + P_{ct}\theta_c k(\tau_t)\mu_{ct}\left[1 - exp(-\alpha_{ct}p_{ct})\right] + \varepsilon_{ct} \tag{14}$$

$$d_{ct} = \frac{\mu_{ct}k(\tau_{ct})}{\chi + exp(\alpha_{ct}p_{ct})} \tag{15}$$

$$p_{ct} = \frac{1}{\alpha_{ct} - 1}\left\{ln\left[\alpha_{ct}k(\tau_{ct})\right] + ln\left[\frac{a_c}{a_c + b}\left(z_{ct} - P_{ct}\Omega_{ct}\overline{\theta}^C_{ct} + 1\right) + (1 - P_{ct})\varrho^{med}_N + P_{ct}\varrho^{med}_P\right] - 1\right\} \tag{16}$$

$$\overline{\theta}^C_{ct} \equiv \frac{1}{\sigma^2_\xi}B_c + \frac{1}{\sigma^2_q}\sum_{s=s_0}^{t-1}\left[q_{cs} - \Theta_S - \mu_{cs}k(\tau_{cs})exp(-\alpha_{cs}p_{cs})\right]\omega_{cs} \tag{17}$$

where ϱ^{med}_j, $j = N, P$ are distributed according to the densities derived in Appendix 2. In equations (14) and (15) the sources of randomness are ε_{ct} and μ_{ct} respectively; on the other hand, equilibrium police enforcement (equation (16)) 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 $\overline{z}_{ct}$ and variance σ^2_z: $z_{ct} \sim N(\overline{z}_{ct}, \sigma^2_z)$. 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 ϱ^i).

Given that the parameters of the model are identified only up to scale, I normalize the variance of individual belief biases ξ^i to 1. Interpretation of all other parameters will thus be relative to ξ^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 $\{\{\Omega_{ct}\overline{\theta}^C_{ct}\}^T_{t=1}\}^N_{c=1}$, and of the shape of the distribution of the median voter's unobserved preferred enforcement type ϱ^{med}_j. While b, σ^2_q, σ^2_z, χ, κ, and λ are assumed constant across cities, I allow the rest of parameters to vary with observable community characteristics. Parameters to be estimated are listed below:

Parameters	
Effect of Prohibition on crime	$\{\theta_c = \mathbf{x}_{\boldsymbol{c}}^{\boldsymbol{\theta}\prime}\boldsymbol{\Lambda}\}_{c=1}^{N}$
Alcohol market size	$\{a_c = \overline{\mathbf{x}}_{\boldsymbol{c}}^{\boldsymbol{a}\prime}\boldsymbol{\Gamma}_a, b\}_{c=1}^{N}$
Law Enforcement	$\{\{\alpha_{ct} = \mathbf{x}_{\boldsymbol{ct}}^{\boldsymbol{e}\prime}\Psi\}_{t=1}^{T}\}_{c=1}^{N}$
Collective Prior	$\{B_c = \mathbf{x}_{\mathbf{c0}}^{\boldsymbol{B}\prime}\boldsymbol{\Xi}\}_{c=1}^{N}$
State-specific crime shifter	$\{\Theta_S = \mathbf{x}_{\boldsymbol{S}}^{\boldsymbol{\Theta}\prime}\boldsymbol{\Sigma}\}_{\forall S}$
Mean moral views	$\{\{\overline{z}_{ct} = \mathbf{x}_{\boldsymbol{ct}}^{\boldsymbol{M}\prime}\boldsymbol{\Pi}\}_{t=1}^{T}\}_{c=1}^{N}$
Arrest probability	χ
Alcohol supply catch-up	κ, λ
Variance of prior moral views	σ_{ζ}^2
Variance of moral externality	σ_{z}^2
$Corr(\zeta^i, \xi^i)$	ρ

$\mathbf{x}_S^{\Theta}$ includes state-level dummies, $\mathbf{x}_{\boldsymbol{c}}^{\boldsymbol{\theta}}$ includes border cities, South, state-capitals indicators, average demographics, and a constant, $\mathbf{x}_{\mathbf{c0}}^{\boldsymbol{B}}$ contains the initial religious ascriptions distribution and a constant, $\overline{\mathbf{x}}_{\boldsymbol{c}}^{\boldsymbol{a}}$ includes average demographics, average religious ascriptions, average population, and a constant, $\mathbf{x}_{\boldsymbol{ct}}^{\boldsymbol{e}}$ is a vector of legal enforcement variables (and a constant) such as the number of state-level dry laws in place, 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), and $\mathbf{x}_{ct}^{M}$ is a vector of containing the religious ascriptions distribution, and a constant [25][26].

6.1 The Likelihood Function

I estimate the equilibrium political-economy model developed above through Conditional Maximum Likelihood (CMLE). Individuals learn about θ_c by observing the realizations of the outcome vector $\boldsymbol{y}_{ct} = (p_{ct}, d_{ct}, q_{ct})$. The system in (14)-(16) has a particularly convenient "triangular" structure, which moreover, justifies the learning process implied by Bayesian learning. Once p_{ct} is realized, conditional on ϱ^{med} individuals face no uncertainty coming from equation (16) (recall that individuals observe z_{ct}). Then, after d_{ct} is realized, the realization of μ_{ct} can be exactly backed-up from equation (15). As a result, in equation (14) the only remaining uncertainty about crime comes from ε_{ct} and

[25] While the first moment of the beta distribution is determined by the relative magnitudes of 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.

The mean of the beta distribution is not identified. Hence, I impose the following normalization: For the city with a value of $\overline{\mu}$ closest to its mean (from equation (2)) across cities in the sample for structural estimation, I set its mean wet share $\frac{a_c}{a_c+b}$ to be equal to the city's value of $\overline{\mu}$. In the sample, this city corresponds to Albany, NY, which has a value of $\overline{\mu} = 0.5107$. This normalization makes the interpretation of the estimated parameters a_c and b more transparent.

[26] The levels of drunkenness arrests for Boston and Chicago are an order of magnitude larger than for the rest of the sample. Thus, I allow for a specific differential level of χ for these two cities.

beliefs about θ_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 $\boldsymbol{y}_{ct} = (p_{ct}, d_{ct}, q_{ct})$. It is the product of a beta density coming from the distibution of the alcohol market size μ_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 $\boldsymbol{y}_{ct}$ 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 θ_c, which depend on the whole sequence of previous realizations of the crime rate during Prohibition years $\{q_{cs}\}_{s=s_0}^{t-1}$.

Let $\boldsymbol{\beta} \equiv (\boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\Xi}, \boldsymbol{\Gamma}_a, b, \chi, \kappa, \lambda, \boldsymbol{\Psi}, \boldsymbol{\Pi}, \sigma_q^2, \sigma_z^2)$, and $\mathbf{x}_{ct} \equiv (\mathbf{x}_{\boldsymbol{S}}^{\boldsymbol{\Theta}}, \mathbf{x}_c^{\theta}, \mathbf{x}_{\boldsymbol{c0}}^{\boldsymbol{B}}, \overline{\mathbf{x}}_{ct}^a, \mathbf{x}_{ct}^M, \mathbf{x}_{c0}^B, \mathbf{x}_{ct}^e)$. The conditional likelihood can be more compactly written as $\mathcal{L}_{ct}(\boldsymbol{y}_{ct}; \boldsymbol{y}_{ct-1}, \mathbf{x}_{ct}, \boldsymbol{\beta} | \varrho_j^{med}, P_{ct}, \tau_{ct})$, 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 $\boldsymbol{y}_{ct}$ only depends on $\boldsymbol{y}_{ct-1}$ through the learning channel, and hence the DGP is dynamically complete (See Wooldrige (2002, p. 412)). As a result, conditional on $\boldsymbol{y}_{ct-1}$, the $\boldsymbol{y}_{ct}$ are independently distributed. Thus, the conditional likelihood for a given observation $\boldsymbol{y}_c = (\boldsymbol{y}_{c1}, \boldsymbol{y}_{c2}, ...\boldsymbol{y}_{cT})'$ is given by $\mathcal{L}_c(\boldsymbol{y}_c, \boldsymbol{\beta} | \varrho^{med}, \boldsymbol{P}_c, \boldsymbol{\tau}_c) = \prod_t \mathcal{L}_{ct}(\boldsymbol{y}_{ct}; \boldsymbol{y}_{ct-1}, \mathbf{x}_{ct}, \boldsymbol{\beta} | \varrho^{med}(P_{ct}), P_{ct}, \tau_{ct})$, where ϱ^{med} is drawn from $f_{\varrho_P^{med}}(\varrho^{med}; a_c, b, \sigma_\zeta^2, \sigma_\xi^2, \rho)$ during Prohibition years, and from $f_{\varrho_N^{med}}(\varrho^{med}; a_c, b, \sigma_\zeta^2)$ during years without Prohibition. Because the ϱ_j^{med} are unobserved, it is necessary to integrate them out from the conditional likelihood, using their derived equilibrium densities[27]. Estimates of $(\boldsymbol{\beta}, \sigma_\zeta^2, \rho)$ are obtained from the following program:

$$max_{\boldsymbol{\beta}, \sigma_\zeta^2, \rho} \sum_c ln \left\{ \int_{-1}^{0} \left[\prod_t \mathcal{L}_{ct}(\boldsymbol{y}_{ct}; \boldsymbol{y}_{ct-1}, \mathbf{x}_{ct}, \boldsymbol{\beta} | \varrho^{med}(P_{ct})) \right] f_{\varrho^{med}(P_{ct})}(\varrho^{med}; a_c, b, \sigma_\zeta^2, \rho, P_{ct}) d\varrho^{med} \right\} \quad (18)$$

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 to identify the alcohol market dynamics, I estimate the structural model for this limited 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.

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

[27] Dynamic models estimated by MLE usually face an "initial conditions" problem since the observation for the first year in the sample depends upon an unobserved realization of the endogenous variable. Here such a problem does not arise because for years under no Prohibition, the likelihood function does not depend on previous realizations of $\boldsymbol{y}_c$. For the first Prohibition year, the learning model implies that beliefs are exclusively based on the prior $\overline{\theta}_{c0}$, which is not a function of $\boldsymbol{y}_{ct-1}$ either. For all subsequent years under Prohibition, the relevant lagged information is available.

in the United States during those years. Thus, for estimation I use the de-trended police expenditure share as the measure for p_{ct}[28]. As the crime outcome measure, I use the natural logarithm of the homicide rate, which standarizes the variance in homicide rates across cities, and is consistent with the shocks in equation (14) being normally distributed, and drawn from the same distribution across cities. See appendix 5 for further discusion of the data).

6.2 Moments Identifying the Parameters in the Model

Here I briefly discuss the relevant moments identifying the different parameters of the model. The structural elasticity of crime to the adoption of Prohibition, θ_c, is a function of city characteristics. 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 θ_c, given that equation (14) 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 χ, 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 χ is assumed constant over time and across cities, χ is identified from the variation in arrest rates that is common across cities over time.

Regarding α_{ct}, the model assumes the dynamics of the legal standard are exogenous to the city. Although citizens were voting both for local law enforcement and for state and federal legislation, the assumption is that within a state or the Country as a whole, each city was too small to affect the equilibrium choice of legal standard. This seems like a natural assumption, given that citizens

[28] While the average annual growth rate of total public spending in the sample is 5.6% (s.e. $=$ 2.2%), it is 3.7% (s.e.=2.5%) for police expenditure. To obtain the detrended police share variable I ran a regression of the raw police expenditure share p^r_{ct} for each city in the sample, on a city-specific linear time-trend, city effects, and no constant: $p^r_{ct} = \alpha_c + \beta_c t + v_{ct}$. I then compute the detrended police share as $p_{ct} = \alpha_c + \widehat{v}_{ct}$. This is equivant to running a separate regression for each city.

in 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 α_{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 observed police enforcement choices and optimal choices under perfect information. Because the model estimates θ_c, 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.

The κ and λ parameters are identified off 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. 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 (6) is very flexible and can accommodate a wide variety of nonlinear trends.

Average moral views $\overline{z}$, which are function of the religious ascription distribution in the community are identified, from equation (16), 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 $\overline{z}$ comes from the variation in law enforcement which is common for the city before and during Prohibition. On the other hand σ_z^2, the second moment of the distribution of z_{ct}, is identified directly from the sample variation in police enforcement that is common across cities.

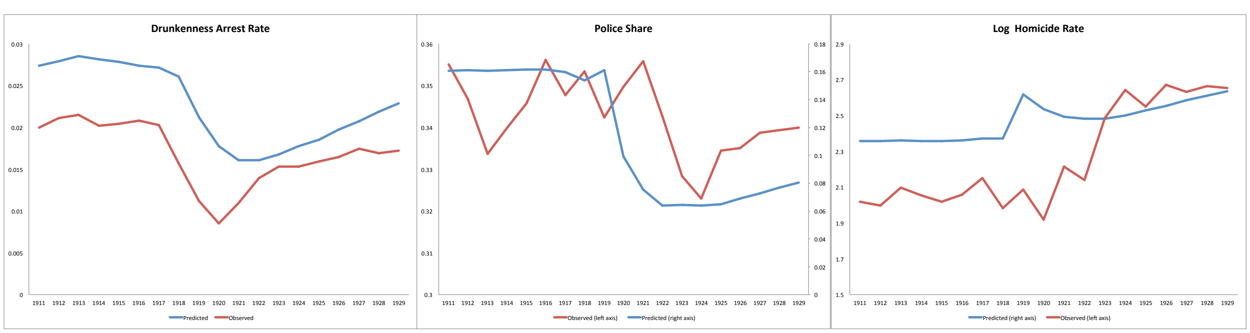
Finally, σ_ζ^2 and ρ are identified in the model from the change in the shape of the estimated density of ϱ^{med} between no-Prohibition and Prohibition years. As the variance of the distribution of idiosincratic moral views decreases, the density of ϱ_N^{med} becomes more bimodal relative to the density of ϱ_P^{med}. This is because individuals become more homogeneous on the moral dimension which, as a result, makes the preference for alcohol consumption more salient in individuals' preferences. On the other hand, as ρ increases in magnitude the differential law enforcement decision of dry cities relative to wet communities is magnifies, increasing the variance of the distribution of ϱ_P^{med} relative to the distribution of ϱ_N^{med}. The reason is that if moral views ζ^i and belief biases ξ^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 ρ increases in magnitude, the density under Prohibition shifts mass to the left, making lower values of police expenditure more likely. Thus, ρ is of special interest in the estimation since it is identified off 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, Figure 5 presents the average (across cities) observed and predicted outcomes, for the sample years. The predictions are computed directly from equations (14)-(17), where I use the estimated expected value for the wet share μ_{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 ϱ^{med}, which I calculate by integrating over the estimated equilibrium densities $f_{\varrho_P^{med}}(\varrho_P^{med}; a_c, b, \sigma_{\varrho Pt})$ and $f_{\varrho_N^{med}}(\varrho_N^{med}; a_c, b, \sigma_{\varrho N})$. 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. On the other hand, the predicted magnitude of the fall in the drunkenness arrest rate closely follows the one observed in the data between 1916 and 1920, except for a difference in the overall level. Finally, the figure shows that the model overpredicts the level of the homicide rate during the 1910s, and also predicts a smaller increase in this variable, compared to the sharp 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.

Figure 5: Fit of the Model



In addition, a way to asses the fit of the model is to look at the variability in the average moral views required to match the data. From equation (16), if the evolution of law enforcement, the alcohol supply, beliefs, and the change in the distribution of ϱ_j^{med} 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 $\sigma_z^2 = 0.31$, which is a third of σ_ξ^2 (normalized to 1). Overall, the estimates suggest that the mechanisms highlighted in the model 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 3, and Table 4 presents the implied average (across cities) estimated values of the main parameters of the model. Standard errors for the coefficients are computed through a bootstrap of size 100. Among the covariates for a_c, most of the coefficients are unprecisely estimated. Population significantly reduces the size of a_c. Together, the average estimate of a_c across cities is 0.93 and is 1.67 for b. Since the variation of a_c across cities is small (its standard deviation is 0.48), the model predicts very similar sizes of the "drinking population" across cities. Nevertheless, the large standard errors associated with the covariates for a_c suggest a weak relationship between the demand for alcohol and the religious distribution.

Looking at the covariates for average moral views $\overline{z}_c$, Orthodox, Evangelical and Mormon shares significantly increase average moral views. Surprisingly, the coefficient for the Catholic share is positive too (1.5), but imprecisely estimated (s.e. $= 0.9$). Looking at the covariates for α_{ct}, the number of alcohol-related laws variable is highly significant and negative (point estimate $= -0.05$, s.e.$= 0.014$), suggesting that changes in dry legislation were actually correlated with a lower effectiveness of policing. This is exactly what a general equilibrium political economy mechanism would suggest, since the passage of state-level dry legislation can be undone by the local level choices of enforcement effectiveness when local authorities are unwilling to enforce Prohibition laws which their communities oppose. On the other hand, the coefficient on the Enforcement Law dummy is negative but insignificant (point estimate $= -0.29$, s.e.$= 0.49$), suggesting the repeal of state-level Prohibition enforcement laws did not alter the effectiveness of policing for crime enforcement. None of the Prohibition Unit indicators show a significant differential effectiveness of enforcement relative to the New York Unit, suggesting similar enforcement technologies across the country.

Of central interest are the model's estimates of θ_c, the structural "elasticity" of Prohibition enforcement to crime. The average θ_c is 1.8, with a standard deviation of 0.7. Among the estimates for its covariates, Table 3 shows that border cities (Canadian or Mexican border or coastal city) had significantly larger elasticities (point estimate $= 0.655$, s.e.$=$ .0.25). These were generally more valuable regions for smuggling networks to control, and at the same time, places where federal enforcement was more intense. Interestingly, the Southern dummy estimate is negative (point estimate $= -1.7$, s.e. $= 0.79$), suggesting that the increase in criminality due to the introduction of Prohibition was differentially lower in the South. At the means of the police share p_{ct} and the estimated parameters, the average city saw an increase in the homicide rate of around 50% during Prohibition[29].

[29] The average (normalized) police share is 0.36. Assuming an 70% size of the alcohol supply (around the 7th year under Prohibition using the estimated κ and λ), and using the mean estimate of $\alpha_{ct} = 4.35$, $a_c/(a_c + b) = 0.35$ and $\theta_c = 1.8$, it follows that $0.52 = exp(1.8 \times 0.7 \times 0.35 \times [1 - exp(-4.35 \times 0.36)]) - 1$.

Table 3: Conditional Maximum Likelihood Estimates

	Covariate	Coefficients
a	% ages 15-44*	2.839
		(5.198)
	% Foreign White*	-1.603
		(9.699)
	% Black*	-1.058
		(4.916)
	% Baptist*	1.593
		(7.999)
	% Orthodox*	-1.037
		(1.697)
	% Evangelical*	-1.443
		(3.343)
	% Jewish*	5.188
		(8.956)
	% Mormon*	1.271
		(0.737)
	% Lutheran*	1.660
		(5.461)
	% Methodist*	1.064
		(8.859)
	% Catholic*	3.688
		(3.799)
	% Presbyterian*	1.596
		(5.662)
	Log of Population*	-0.479
		(0.086)
	Constant	3.251
		(8.616)
b		1.670
		(0.492)

	Covariate	Coefficients
χ		7.303
		(3.834)
κ		0.236
		(0.009)
λ		0.301
		(0.014)
α	Constant	4.094
		(1.084)
	Number of Alcohol-Related Laws	-0.052
		(0.014)
	Enforcement Law	-0.296
		(0.490)
	Prohibition Unit Seat	-0.021
		(0.670)
	Prohibition Unit: Providence	-0.335
		(1.410)
	Prohibition Unit: Washington	-1.447
		(0.911)
	Prohibition Unit: Jacksonville	-1.282
		(0.728)
	Prohibition Unit: Detroit	-1.307
		(1.119)
	Prohibition Unit: Chicago	0.996
		(4.471)
	Prohibition Unit: Kansas City	-0.256
		(1.220)
	Prohibition Unit: San Francisco	0.605
		(2.611)
	Prohibition Unit: Los Angeles	1.607
		(2.789)
	Prohibition Unit: Seattle	2.854
		(1.603)

	Covariate	Coefficients
z	% Baptist	-0.100
		(0.081)
	% Orthodox	0.118
		(0.041)
	% Evangelical	0.049
		(0.018)
	% Jewish	0.246
		(0.293)
	% Mormon	0.334
		(0.018)
	% Lutheran	0.105
		(0.105)
	% Methodist	-0.060
		(0.125)
	% Catholic	1.507
		(0.929)
	% Presbyterian	-0.008
		(0.071)
	Constant	2.370
		(1.468)
θ	Border	0.655
		(0.252)
	South	-1.732
		(0.791)
	State Capital	-0.040
		(0.173)
	Share Ages 15-44*	0.591
		(0.796)
	Share Foreign White*	0.417
		(1.064)
	Share Black*	0.229
		(1.328)
	Constant	1.425
		(0.621)

	Covariate	Coefficients
B	% Baptist in 1911	0.148
		(0.136)
	% Orthodox in 1911	-0.004
		(0.015)
	% Evangelical in 1911	-0.175
		(0.010)
	% Jewish in 1911	-0.062
		(0.004)
	% Mormon in 1911	-0.421
		(0.037)
	% Lutheran in 1911	-0.124
		(0.091)
	% Methodist in 1911	0.376
		(0.110)
	% Catholic in 1911	-0.552
		(0.389)
	% Presbyterian in 1911	0.102
		(0.052)
	Constant	-0.723
		(0.635)
σ_q^2		0.253
		(0.009)
σ_z^2		0.314
		(0.416)
σ_ξ^2		1.000
		(0.000)
σ_ζ^2		2.348
		(0.495)
ρ		-0.489
		(0.002)

Log-likelihood:	4582.20
Observations:	1254

*1911-1929 averages

Note: Standard Errors computed through a bootstrap of size 100.

Estimates of the State Effects are omitted from the table.

Estimates for the covariates of the Prior B_c are also presented in Table 3. Evangelical and Mormon shares are significantly correlated with more negative biases (more initial optimism), as expected. In contrast, the estimate for the Methodist share is actually positive. On the other hand, the estimates for the second moments of the joint distribution of individual biases and moral views (see equation (4)) also are of interest. The variance of individual moral views σ_ζ^2 is estimated to be 2.34 (s.e. = 0.49), implying that variation in individual moral views was significantly larger than variation in biases (normalized to 1). Thus, across the population there was much more initial agreement on the effects of Prohibition than on its moral virtues. Finally, ρ, the estimated correlation between prior biases (ξ^i) and moral views (ζ^i) is -0.48 (s.e.= 0.002), suggesting that cities with constituencies more favorable to Prohibition did have much more optimistic beliefs about its effects.

Table 4: Average Estimated Values of Parameter Estimates

Parameter	Mean Estimate	Parameter	Mean Estimate
a	0.93	α	4.35
	(0.480)		(1.020)
b	1.670	**z**	3.11
	(0.000)		(0.310)
χ	7.303	θ	1.8
	(0.000)		(0.740)
κ	0.236	**B**	-0.94
	(0.000)		(0.150)
λ	0.301		
	(0.000)		

Note: Standard Deviations in parenthesis

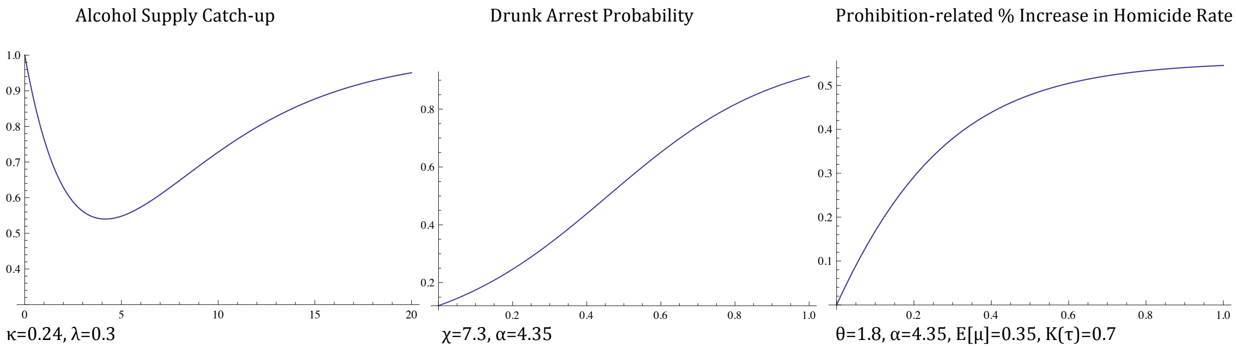
An alternative way to see the correlation between moral views and beliefs from the model's estimates is to regress the estimated average moral views $\overline{z}_c$ on the estimated values of priors B_c for the cities in the sample. The slope of this regression is -1.89 with a t-statistic of -21.55. 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 decrease in policing in the early years under Prohibition (see Figure 6), which in the model is driven by optimistic priors.

The parameter estimates from Table 4 also allow a quantitative characterization of the structural relationships specified in the model. In particular, the estimates for κ and λ from equation (6) are very precisely estimated (0.23, s.e. = 0.009 and 0.3, s.e. = 0.014, respectively), and imply that at its lowest point, the supply of alcohol was on average 55% its pre-Prohibition level, and that this minimum was attained around 3.75 years after the introduction of Prohibition[30]. Together with this estimated function for the alcohol supply catch up, Figure 6 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[31].

[30] The minumim of equation 6 is attained at $\tau = 1/\kappa$.

[31] Using the average parameter estimates from Table 4, the estimated arrest probability is computed as $Pr(Arrest|p) = \frac{exp(4.35p)}{7.3+exp(4.35p)}$, and the estimated proportional increase in the homicide rate under Prohibition is computed as $\Delta Q(p) = exp(1.8 \times 0.7 \times 0.35 \times [1 - exp(-4.35p)])$, for $k(\tau) = 0.7$.

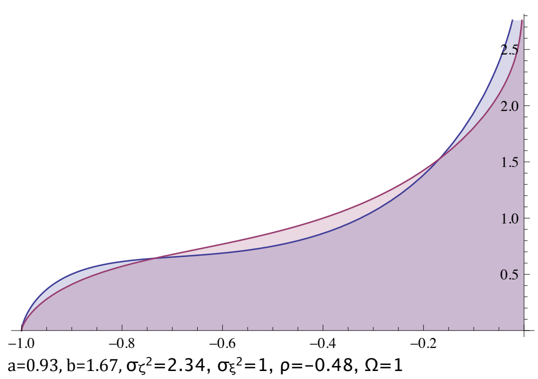
Figure 6: Estimated Functional Forms



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 45%, but only for a relatively short period of time. While increasing policing would increase arrests for drunkenness, the slope is not very steep. A whole standard deviation increase in the police share would at most increase the arrest probability by 4%. In sharp contrast, the same increase in policing during Prohibition would imply that the homicide rate would move from being 41 to 44% higher under Prohibition[32].

The estimated shapes of the distributions of the unobserved ϱ_j^{med} can also be directly derived from the parameter estimates of the structural model, by plugging the estimates of a_c, b, σ_ξ^2, σ_ζ^2, and ρ in equation (24) from Appendix 2. Figure 7 plots both densities (Pink: Prohibition, Light Blue: No Prohibition), for the mean values of the parameter estimates, and for the first year under Prohibition (when $\Omega_t = \sigma_\xi^2$) . The difference in the distributions' shapes under Prohibition and under no Prohibition is what identifies ρ 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.

Figure 7: Estimated Densities of the Median Voters' Unobserved ϱ^{med}

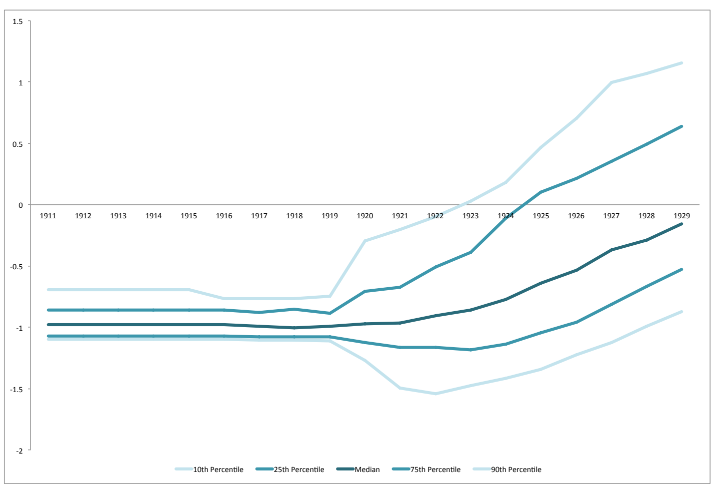


[32] Average (normalized) police share is $p = 0.36$. Thus, the increase in the arrest probability induced from increasing policing to $0.4 = 0.36 + 0.04$ would be $\frac{exp(4.35\times0.4)}{7.3+exp(4.35\times0.4)} - \frac{exp(4.35\times0.36)}{7.3+exp(4.35\times0.36)} = 0.04$. The shift in the homicide rate goes from $exp(1.8\times0.7\times0.35\times[1-exp(-4.35\times0.36)])-1 = 0.41$ to $exp(1.8\times0.7\times0.35\times[1-exp(-4.35\times0.4)])-1 = 0.44$ under Prohibition, when policing is increased by one standard deviation around its average.

6.3.2 Learning

Here I discuss the estimation results related to learning. Differences across cities in the estimated speed of learning are due directly to the variation in enforcement choices over timeUnder normal updating, these affect the informativeness of the signals. There is substantial learning over the nineteen year period. Figure 8 graphs the evolution of the estimated empirical distribution of the common component of beliefs $\{\{\Omega_{ct}\overline{\theta}_{ct}^{C}\}_{c=1}^{N}\}_{t=1911}^{1929}$, derived directly from applying equation (17) 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. Beliefs remain at the prior until cities fall under Prohibition status. For some of the most optimistic cities, early during Prohibition beliefs about θ_c actually fall. After around 1923 though, 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 about Prohibition, the dispersion of beliefs actually increases over time. Mean common beliefs increase from the average prior $B_c = -0.94$ to a mean posterior of -0.001 in 1929, whereas the posterior median is only around -0.15. While the standard deviation of priors is 0.15, it is 0.78 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 have different structural values of θ_c.

Figure 8: Estimated Belief Sequences: Empirical Distribution



A key question is whether the differential evolution of beliefs across cities is correlated with their moral profiles. The reduced-form estimates already suggested that this is the case. In Online Appendix 4 I show that during early Prohibition years wetter cities had differentially lower levels of police enforcement. I argued there that this could be driven by the willingness of more optimistic "dry" cities to invest in law enforcement. 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 2.14[33]. Although the standard deviation of beliefs across cities increased over time, incentives for

[33]For the 66 cities in the sample, I run the regression $\Omega_{c,1929}\overline{\theta}_{c,1929}^{C} = \beta_0 + \beta_1\overline{z}_c + \beta_2\Omega_{c,1911}\overline{\theta}_{c,1911}^{C} + \varepsilon_c$. The estimated

differentially higher initial law enforcement in drier cities limited the extent of beliefs divergence. This is consistent with the fact that among the subset of relatively "wetter" communities, dry ones saw larger shifts of public opinion against Prohibition (see Figure 4b). Overall, the structural estimates 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 (Buera et al. (2011)). Formally, this is equivalent to imposing the restriction $\sigma_\xi^2 = 0$, so that individuals never update their priors. A Likelihood Ratio test compares the restricted No-Learning model with the benchmark model. The log-likelihood for the model without learning is $3,992$, while the log-likelihood for the benchmark model is $4,5822$. Under the null hypothesis that the restricted and unrestricted models are indistinguishable,

$$LR = 2[logL(Benchmark) - logL(NoLearning)] \sim \chi^2_{767} \tag{19}$$

Assuming $\sigma_\xi^2 = 0$ 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 χ^2 distribution is 767. While $LR = 1,179$, the 99% critical value is 861.04. 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 allow a further assessment of the model, and provide general-equilibrium answers to questions of interest, impossible to make in a partial equilibrium or reduced-form framework.

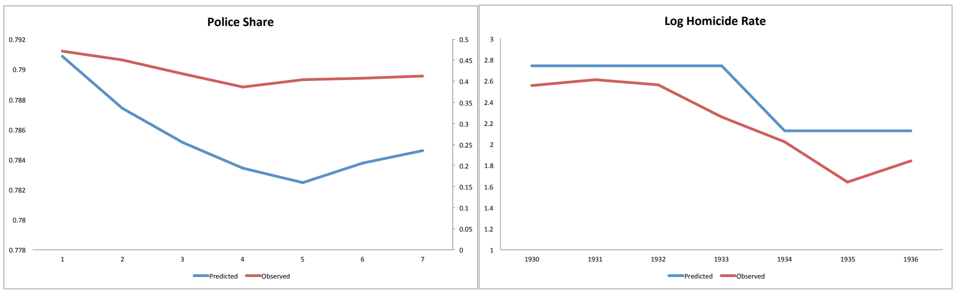
6.4.1 Out of Sample Prediction

Unavailability of drunkenness arrests data for years after 1929 makes me 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, using the MLE estimates on equations (14)-(17). This exercise is particularly meaningful because I do observe the police and homicide rate outcomes in that period, so it assesses the extent to which the model can capture the subsequent evolution of outcomes during Prohibition's final phase. I use the estimated 1929 posterior beliefs as the 1930 priors. I then compute iteratively the predicted equilibrium values of p_{ct} from equation (16), and with them I then predict q_{ct} from equation (14). To compute year t's posterior from equation (17), I add a random shock drawn from a mean-zero normal distribution with variance equal to 0.253 (the MLE estimate for the variance of ε, σ_q^2) 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 stops after this year. Figure 9 presents graphs analogous to those

β_1 is 1.57 with a standard error of 0.73. I include the prior as a regressor to control for the fact that morally drier cities had more negative priors.

in Figure 5, comparing the "out of sample" average predicted values from the structural model over times. The out-of-sample prediction 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.

Figure 9: "Out of Sample" predictions for the years 1930-1936



6.4.2 Changes in Prior Beliefs

The adoption of Prohibition would not have been possible based exclusively on moral motivations, since radically dry sectors did not constitute a large enough majority. Its adoption required a large fraction of morally-indifferent voters with optimistic beliefs about the effects of the policy. What was the cost of these biased prior beliefs? I provide an answer to this question by making the counterfactual exercise of assuming that prior common beliefs in 1911 were unbiased ($B_c = \theta_c$). Using the estimated coefficients, I can 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 and expected homicide outcomes are close to each other given the law enforcement choices. Police enforcement decisions, on the other hand, behave differently. In particular, policing choices would have fallen sharply 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 more aggressive 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 θ_c is larger than the variation in estimated priors. In addition, the model also suggests unbiased priors would have made little difference to the homicide rate. 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. 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. Here I perform a simple exercise, by asking about the evolution of outcomes during Prohibition under a higher degree of polarization in society. By polarization I mean an increase in the average willingness to enforce prohibition, by raising $\overline{z}_{ct}$, coupled with an increase in alcohol demand, by raising the mean of the distribution of μ. Thus, I allow $\overline{z}_{ct}$, but also $E[\mu]$, to increase by one or two standard deviations. The estimated standard deviation of the "wet share" μ is 0.25, while its mean is 0.36, so that a one standard deviation increase in the mean implies $E[\mu] = 0.6$. Holding a fixed, such a shift in the distribution of the wet share can be achieved by reducing the value of b to 0.62. For a two standard deviation increase in the mean of μ, which implies $E[\mu] = 0.85$, a value of $b = 0.22$ achieves the same objective. The model predicts that the speed of learning during Prohibition increases very fast on the degree of polarization in society, but most of the learning takes place in the first 4 to 5 years under Prohibition. Subsequently learning slows down. The benchmark model's estimated 1929 posterior 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. 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.

On the other hand, policing choices would have been much more stable over time because the increased salience of the moral externality reduces the extent to which 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 3.7 more homicides per hundred thousand on the average Prohibition year in the two standard deviations higher polarization society. Thus, although communities with more extreme preference distributions do learn much faster, they also face a constituency that is more willing to endure the increased levels of crime.

6.4.4 Alternative Political Environments

It is natural to ask about the equilibrium effects of changes in the political environment. As I have shown, equilibrium 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 $\varrho_j = -1$), or the median voter among the drys (the decisive voter's type is $\varrho_j = 0$). Under each conterfactual scenario I compute the predicted outcome sequences, using the benchmark parameter estimates.

When drys have all the political power, law enforcement chioces are consistently higher 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 are on avergage two years ahead relative to the benchmark case. Consistently, the predictions of the counterfactual simulation where wets have all the political power deliver weaker law enforcement sequences, which translate 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 and average voter's preferences. They highlight that the effects of increased conflict also arise when the identity of those deciding over law enforcement diverges 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 (wets) have all the political power at the local level, their law enforcement choices are too high (low) relative to what the community's median voter would prefer. As a result, crime outcomes are more (less) informative and communities learn faster (more slowly).

7 Conclusions

Many central political cleavages in contemporary societies revolve around ideological or moral issues, over which people often 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 Twentieth century. The circumstances around it were very specific. The potential effects over crime from closing the alcohol market 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. 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 results suggest that an key element to understand the effects and success of policies is the degree of alignment between 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. Estimates suggest 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 were responsive to learning. The model's assumption of exclusively localized learning appears consistent with the data. In other contexts policy learning spillovers might prove important. For example, neighboring communities might be a key source of infromation in societies where the media plays a large role in shaping public opinion. This constitutes an avenue for understanding other instances of social change. This paper did not exploit the judicial dimension of law enforcement, 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

Alesina, Alberto and Nicola Fuchs-Schundeln, "Good-Bye Lenin (or Not?) The Effect of Communism on People's Preferences," *American Economic Review*, 2007, *97* (4), 1507–1528.

Asbury, Herbert, *The Great Illusion. An Informal History of Prohibition*, Garden City, NY: Doubleday and Co., 1950.

Becker, Gary, "Crime and Punishment," *The Journal of Political Economy*, March 1968, *76* (2), 169–217.

Blocker, Jack, *American Temperance Movements. Cycles of Reform*, Boston, MA: Twayne Publishers, 1989.

_ , "Did Prohibition Really Work? Alcohol Prohibition as a Public Health Innovation," *Public Health Then and Now*, 2006, *96* (2), 233–243.

Buera, Francisco, Alexander Mongue-Naranjo, and Giorgio Primiceri, "Learning the Weath of Nations," *Econometrica*, January 2011, *79* (1), 1–45. Unpublished.

Coate, Stephen and Michael Conlin, "A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence," *American Economic Review*, 2004, *94* (5), 1476–1504.

Colvin, Leigh, *Prohibition in the United States*, New York: George H. Doran Company, 1926.

Degan, Arianna and Antonio Merlo, "A Structural Model of Turnout and Voting in Multiple Elections," March 2009.

Dills, Angela, Jeffrey Miron, and Garrett Summers, "What do Economists Know About Crime?," January 2008. NBER Working Paper No. 13759.

Donohue, John and Steve Levitt, "The Impact of Legalized Abortion on Crime," *The Quarterly Journal of Economics*, 2001, *116* (2), 379–420.

Foster, Gaines, *Moral Reconstruction: Christian Lobbyists and the Federal Legislation of Morality, 1865-1920*, The University of Carolina Press, 2002.

Franklin, Jimmie Lewis, *Born Sober: Prohibition in Oklahoma, 1907-1959*, Norman, OK: University of Oklahoma Press, 1971.

Goldstein, Paul, "The Drugs/Violence Nexus: A Tripartite Conceptual Framework," *Journal of Drug Issues*, 1985, *39*, 143–174.

Haider-Markel, Donald and Kenneth Meier, "The Politics of Gay and Lesbian Rights: Expanding the Scope of the Conflict," *The Journal of Politics*, May 1996, *58* (2), 332–349.

Harrison, Leonard and Elizabeth Laine, *After Repeal. A Study of Liquor Control Administration*, New York: Harper and Brothers, 1936.

Hayek, Frederick, *The Constitution of Liberty*, Chicago, IL: The Chicago University Press, 1960.

Isaac, Paul, *Prohibition and Politics: Turbulent Decades in Tennessee 1885-1920*, Knoxville, TN: The University of Tennessee Press, 1965.

Kyvig, David, *Repealing National Prohibition*, Chicago: The University of Chicago Press, 1979.

Lagunoff, Roger, "A Theory of Constitutional Standards and Civil Liberties," *Review of Economic Studes*, 2001, *68* (1), 109–132.

Landier, Augustin, David Thesmar, and Mathias Thoening, "Investigating Capitalism Aversion," *Economic Policy*, 2008, *23* (55), 465–497.

Levitt, Steven, "Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not," *The Journal of Economic Perspectives*, 2004, *18* (1), 163–190.

Miron, Jeffrey, "The Effect of Alcohol Prohibition on Alcohol Consumption," 1999. NBER Working Paper No. 7130.

Mukand, Sharun and Dani Rodrik, "In Search of the Holy Grail: Policy Convergence, Experimentation and Economic Performance," *American Economic Review*, 2005, *95* (1), 374–383.

of Prohibition, U.S. Bureau, "Statistics Concerning Intoxicating Liquors," Several Years.

Okrent, Daniel, *Great Fortune: The Epic of Rockefeller Center*, New York: Viking Press, 2003.

_, *Last Call: The Rise and Fall of Prohibition*, New York: Scribner, 2010.

Owens, Emily, "Are Underground Markets Really More Violent? Evidence from Early 20th Century America," *American Law and Economics Review*, 2011, *13* (1), 1–44.

Schmeckebier, Laurence, *The Bureau of Prohibition. Its History, Activities and Organization*, Washington D.C: The Brookings Institution, 1929.

Sinclair, Andrew, *Prohibition: The Era of Excess*, London: Faber and Faber.

Smith, Alfred, *The New York Red Book*, Vol. 1923, Albany, NY: J. B. Lyon Co., 1923.

Stayton, W. H., "Our Experiment in National Prohibition. What Progress Has it Made?," *Annals of the American Academy of Political and Social Science*, 1923, *109*, 26–38.

Strulovici, Bruno, "Learning While Voting; Determinants of Collective Experimentation," *Econometrica*, 2010, *78* (3), 933–971.

Szymansky, Anne-Marie, *Pathways to Prohibition*, Durham, NC: Duke University Press, 2003.

Tydings, Millard E., *Before and After Prohibition*, New York: The Macmillan Company, 1930.

Vives, Xavier, *Information and Learning in Markets: The Impact of Market Microstructure*, Princeton, NJ: Princeton University Press, 2010.

Wickersham-Commission, "Wickersham Commission Records, 1928-1931," Original Papers and Documents, Harvard University Law School Library 1928-1931.

Wooldrige, Jeffrey, *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2002.

Appendix 1: Derivation of Ideal Law Enforcement Choice

Indirect utility is given by

$$E_tU^i(p_t|P_t=1) = 1_{\{i\in W_t\}}k(\tau_t)exp(-\alpha_t p_t) - z^i\frac{a}{a+b}k(\tau_t)exp(-\alpha_t p_t)$$

$$+\,exp(1-p_t) - \Theta_S - \frac{a}{a+b}k(\tau_t)exp(-\alpha_t p_t) - P_t\overline{\theta}_t^i k(\tau_t)\frac{a}{a+b}[1-exp(-\alpha_t p_t)] \tag{20}$$

The first order condition with respect to p_{ct} from equation (20) is

$$-1_{\{i\in W_{ct}\}}k(\tau_{ct})\alpha_{ct}exp(-\alpha_{ct}p_{ct}) + \alpha_{ct}\frac{a_c}{a_c+b}k(\tau_{ct})z_c^i exp(-\alpha_{ct}p_{ct}) - exp(1-p_{ct})$$

$$+\alpha_{ct}\frac{a_c}{a_c+b}k(\tau_{ct})exp(-\alpha_{ct}p_{ct}) - P_t\alpha_{ct}\overline{\theta}_{ct}^i\frac{a_c}{a_c+b}k(\tau_{ct})exp(-\alpha_{ct}p_{ct}) \leq 0$$

Solving for p_{ct}, equation (12) directly follows. The second-order condition for the solution in equation (12) to be a maximum is

$$\Leftrightarrow 2ln\alpha_{ct} + ln\left[\frac{a_c}{a_c+b}(z_c^i - \overline{\theta}_{ct}^i + 1) - 1_{\{i\in W_t\}}\right] - 1 > (\alpha_{ct}-1)p_{ct} \tag{21}$$

I verify this condition is satisfied for the parameter estimates.

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 (ζ^i, ξ^i), 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 $\varrho_{DP}^i \equiv \frac{a}{a+b}(\zeta^i - \Omega_t\frac{1}{\sigma_\xi^2}\xi^i)$ (DP for Dry under Prohibition), and $\varrho_{WP}^i \equiv \frac{a}{a+b}(\zeta^i - \Omega_t\frac{1}{\sigma_\xi^2}\xi^i) - 1$ (WP for Wet under Prohibition). These are normal random variables distributed according to $\varrho_{DP}^i \sim N(0, \sigma_{\varrho Pt}^2)$ and $\varrho_{WP}^i \sim N(-1, \sigma_{\varrho Pt}^2)$ respectively, where $\sigma_{\varrho Pt}^2 \equiv \left(\frac{a}{a+b}\right)^2\left(\sigma_\zeta^2 + \Omega_t^2\frac{1}{\sigma_\xi^2} - 2\Omega_t\rho\frac{\sigma_\zeta}{\sigma_\xi}\right)$[34]. Now define $\varrho_P^i \equiv 1_{\{i\in D_t\}}\varrho_{DP}^i + 1_{\{i\in W_t\}}\varrho_{WP}^i$. The

[34] This variance is time-varying. As learning takes place and $\Omega_t \to 0$, $\sigma_{\varrho Pt}^2 \to \sigma_\zeta^2$.

conditional density of ϱ_P^i is given by

$$f_{\varrho P}(\varrho_P^i|\mu_t) = (1-\mu_t)N(0,\sigma_{\varrho Pt}^2) + \mu_t N(-1,\sigma_{\varrho Pt}^2)$$

since with probability μ_t a wet individual is sampled, and with probability $1-\mu_t$ a dry individual is sampled. Thus, the distribution of ϱ_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 ϱ_{WP}^i and ϱ_{DP}^i, as $\mu_t \to 0$, the median voter's type $\varrho_P^{med} \to 0$, and as $\mu_t \to 1$, $\varrho_P^{med} \to -1$, so that $\varrho_P^{med} \in (-1,0)$. For periods under no Prohibition, analogously define $\varrho_{DN}^i \equiv \frac{a}{a+b}\zeta^i$ (DN for Dry under no Prohibition) and $\varrho_{WN}^i \equiv \frac{a}{a+b}\zeta^i - 1$ (WN for Wet under no Prohibition), which are distributed according to $\varrho_{DN}^i \sim N(0,\sigma_{\varrho N}^2)$ and $\varrho_{WN}^i \sim N(-1,\sigma_{\varrho N}^2)$ respectively, with $\sigma_{\varrho N}^2 \equiv \left(\frac{a}{a+b}\right)^2 \sigma_\zeta^2$. Now define $\varrho_N^i \equiv 1_{\{i\in D_t\}}\varrho_{DN}^i + 1_{\{i\in W_t\}}\varrho_{WN}^i$, which is a random variable whose conditional density is given by

$$f_{\varrho N}(\varrho_N^i|\mu_t) = (1-\mu_t)N(0,\sigma_{\varrho N}^2) + \mu_t N(-1,\sigma_{\varrho N}^2)$$

Indirect preferences over law enforcement in (12) can be expressed in terms of ϱ_N^i and ϱ_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 ϱ_j^i, so the Median Voter Theorem holds, and the unique political equilibrium value of p_t is given by the preferred choice of law enforcement of the median over the distribution of ϱ_j^i, conditional on μ_t.

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

$$(1-\mu_t)\int_{-\infty}^{\varrho_P^{med}(\mu_t)} \frac{1}{\sqrt{2\pi}\sigma_{\varrho P}} exp\left(-\frac{1}{2\sigma_{\varrho P}^2}\varrho^2\right) d\varrho + \mu_t \int_{-\infty}^{\varrho_P^{med}(\mu_t)} \frac{1}{\sqrt{2\pi}\sigma_{\varrho P}} exp\left(-\frac{1}{2\sigma_{\varrho P}^2}(\varrho+1)^2\right) d\varrho = \frac{1}{2} \quad (22)$$

where I have made explicit the dependence of ϱ_P^{med} on the wet share in the community. Because the realization of μ_t is unobserved, the median ϱ_P^{med} in the population as defined in (22) is a random variable whose density is derived below. The equation analogous to (22) implicitly defining ϱ_N^{med} (the conditional median of the distribution of ϱ_N^i) and its corresponding density are found analogously[35].

Derivation of the density of ϱ_P^{med}:

First, recall that $f_\mu(\mu;a,b)$, the density of μ_{ct}, is beta with parameters (a_c, b). From (22), μ_{ct} can be directly expressed as a function of ϱ_P^{med}:

$$\mu_{ct} \equiv h_\mu(\varrho_P^{med}) = \frac{\frac{1}{2} - \Phi\left(\frac{\varrho_P^{med}}{\sigma_{\varrho Pt}}\right)}{\Phi\left(\frac{\varrho_P^{med}+1}{\sigma_{\varrho Pt}}\right) - \Phi\left(\frac{\varrho_P^{med}}{\sigma_{\varrho Pt}}\right)} \quad (23)$$

[35] Notice that $\sigma_{\varrho Pt}^2 \to \sigma_{\varrho N}^2$ as $\Omega_t \to 0$, which implies that $\varrho_P^{med} \to_d \varrho_N^{med}$.

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

$$f_{\varrho_P^{med}}(\varrho_P^{med}; a_c, b_c, \sigma_{\varrho Pt}) = f_\mu(h_\mu(\varrho_P^{med}); a_c, b_c) \left| \frac{\partial h_\mu(\varrho_P^{med})}{\partial \varrho_P^{med}} \right|$$

The derivative of h_μ is given by

$$\frac{\partial h_\mu(\varrho_P^{med})}{\partial \varrho_P^{med}} = \frac{\frac{1}{\sigma_{\varrho Pt}}\phi\left(\frac{\varrho_P^{med}}{\sigma_{\varrho Pt}}\right)\left[\frac{1}{2} - \Phi\left(\frac{\varrho_P^{med}+1}{\sigma_{\varrho Pt}}\right)\right] - \frac{1}{\sigma_{\varrho Pt}}\phi\left(\frac{\varrho_P^{med}+1}{\sigma_{\varrho Pt}}\right)\left[\frac{1}{2} - \Phi\left(\frac{\varrho_P^{med}}{\sigma_{\varrho Pt}}\right)\right]}{\left[\Phi\left(\frac{\varrho_P^{med}+1}{\sigma_{\varrho Pt}}\right) - \Phi\left(\frac{\varrho_P^{med}}{\sigma_{\varrho Pt}}\right)\right]^2} < 0$$

To see that $\frac{\partial h_\mu(\varrho_P^{med})}{\partial \varrho_P^{med}} < 0$ notice that the first term in square brackets is always smaller than the second term in square brackets. For $\varrho_P^{med} \geq 0$, $\phi\left(\frac{\varrho_P^{med}}{\sigma_{\varrho Pt}}\right) \geq \phi\left(\frac{\varrho_P^{med}+1}{\sigma_{\varrho Pt}}\right)$, and the first term in brackets is more negative than the second term in brackets, so the numerator is negative. For $\varrho_P^{med} < -\frac{1}{2}$, $\phi\left(\frac{\varrho_P^{med}}{\sigma_{\varrho Pt}}\right) < \phi\left(\frac{\varrho_P^{med}+1}{\sigma_{\varrho Pt}}\right)$, 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 $\varrho_P^{med} \in (-\frac{1}{2}, 0)$, $\phi\left(\frac{\varrho_P^{med}}{\sigma_{\varrho Pt}}\right) \geq \phi\left(\frac{\varrho_P^{med}+1}{\sigma_{\varrho Pt}}\right)$, the first term in brackets is negative, and the second term in brackets is positive, so the numerator is negative. Thus, h_μ is a one-to-one mapping, and the density for ϱ_P^{med} is

$$f_{\varrho_P^{med}}(\varrho_P^{med}; a_c, b, \sigma_{\varrho Pt}) = \frac{1}{\sigma_{\varrho Pt}} \frac{1}{\int_0^1 v^{a_c-1}(1-v)^{b-1}dv} \times$$

$$\frac{\phi\left(\frac{\varrho^{med}+1}{\sigma_{\varrho Pt}}\right)\left[\frac{1}{2} - \Phi\left(\frac{\varrho^{med}}{\sigma_{\varrho Pt}}\right)\right]^{a_c}\left[\Phi\left(\frac{\varrho^{med}+1}{\sigma_{\varrho Pt}}\right) - \frac{1}{2}\right]^{b-1} + \phi\left(\frac{\varrho^{med}}{\sigma_{\varrho Pt}}\right)\left[\frac{1}{2} - \Phi\left(\frac{\varrho^{med}}{\sigma_{\varrho Pt}}\right)\right]^{a_c-1}\left[\Phi\left(\frac{\varrho^{med}+1}{\sigma_{\varrho Pt}}\right) - \frac{1}{2}\right]^{b}}{\left[\Phi\left(\frac{\varrho^{med}+1}{\sigma_{\varrho Pt}}\right) - \Phi\left(\frac{\varrho^{med}}{\sigma_{\varrho Pt}}\right)\right]^{a_c+b}} \tag{24}$$

for $\varrho^{med} \in (-1, 0)$, and where $\sigma_{\varrho Pt} = \frac{a_c}{a_c+b}\sqrt{\sigma_\zeta^2 + \Omega_{ct}^2 \frac{1}{\sigma_\xi^2} - 2\Omega_{ct}\rho\frac{\sigma_\zeta}{\sigma_\xi}}$. Replacing $\sigma_{\varrho N} = \frac{a_c}{a_c+b}\sigma_\zeta$ for $\sigma_{\varrho Pt}$ everywhere in (24), the density of unobserved heterogeneity in preferred law enforcement during periods under no Prohibition is obtained: $f_{\varrho_N^{med}}(\varrho_N^{med}; a_c, b_c, \sigma_{\varrho N})$. $\square$

Appendix 3: Derivation of the Conditional Likelihood

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

$$f_{z\mu\varepsilon}(z_{ct}, \mu_{ct}, \varepsilon_{ct}; a_c, b, \overline{z}_{ct}, k, \lambda, \sigma_q^2, \sigma_z^2) = \frac{1}{\sqrt{2\pi}\sigma_z} exp\left(-\frac{1}{2\sigma_z^2}(z_{ct} - \overline{z}_{ct})^2\right) \frac{\mu_{ct}^{a_c-1}(1-\mu_{ct})^{b-1}}{\int x^{a_c-1}(1-x)^{b-1}dx} \frac{1}{\sqrt{2\pi}\sigma_q} exp\left(-\frac{\varepsilon_{ct}^2}{2\sigma_q^2}\right)$$

From (14), (15), and (16), $(z_{ct}, \mu_{ct}, \varepsilon_{ct})$ can be expressed as a function of the observables (p_{ct}, d_{ct}, q_{ct}):
From (16),

$$z_{ct} \equiv g_z(p_{ct}, d_{ct}, q_{ct}; \varrho_N^{med}, \varrho_P^{med}) = \frac{a_c+b}{a_c}\frac{1}{\alpha_{ct}k(\tau_{ct})} exp((\alpha_{ct}-1)p_{ct}+1) - \frac{a_c+b}{a_c}\left[P_{ct}\varrho_P^{med} + (1-P_{ct})\varrho_N^{med}\right] + P_{ct}\Omega_{ct}\overline{\theta}_{ct}^C - 1$$

From (15),

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

Finally from (14), and replacing for μ_{ct} from above,

$$\varepsilon_{ct} \equiv g_\varepsilon(p_{ct}, d_{ct}, q_{ct}) = q_{ct} - \Theta_S - d_{ct}(\chi + exp(\alpha_{ct}p_{ct}))\left\{exp(-\alpha_{ct}p_{ct}) + P_{ct}\theta_c\left[1 - exp(-\alpha_{ct}p_{ct})\right]\right\}$$

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

$$f_{pdq}(p_{ct}, d_{ct}, q_{ct}) = f_{z\mu\varepsilon}(g_z(p_{ct}, d_{ct}, q_{ct}; \varrho_N^{med}, \varrho_P^{med}), g_\mu(p_{ct}, d_{ct}, q_{ct}), g_\varepsilon(p_{ct}, d_{ct}, q_{ct}); a_c, b, \overline{z}_{ct}, k, \lambda, \sigma_q^2, \sigma_z^2)\,|J_{ct}|$$

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

$$|J_{ct}| = \begin{vmatrix} \frac{\partial g_z}{\partial p_{ct}} & \frac{\partial g_z}{\partial d_{ct}} & \frac{\partial g_z}{\partial q_{ct}} \\ \frac{\partial g_\mu}{\partial p_{ct}} & \frac{\partial g_\mu}{\partial d_{ct}} & \frac{\partial g_\mu}{\partial q_{ct}} \\ \frac{\partial g_\varepsilon}{\partial p_{ct}} & \frac{\partial g_\varepsilon}{\partial d_{ct}} & \frac{\partial g_\varepsilon}{\partial q_{ct}} \end{vmatrix}$$

Given the structure of the model, conveniently $\frac{\partial g_z}{\partial d} = \frac{\partial g_z}{\partial q} = \frac{\partial g_\mu}{\partial q} = 0$, and $\frac{\partial g_\varepsilon}{\partial q} = 1$. To show that $g(p_{ct}, d_{ct}, q_{ct})$ is a one-to-one mapping, it is sufficient that $\frac{\partial g_z}{\partial p}$, $\frac{\partial g_\mu}{\partial d}$, $\frac{\partial g_\mu}{\partial p}$, $\frac{\partial g_\varepsilon}{\partial d}$, and $\frac{\partial g_\varepsilon}{\partial p}$ do not change sign. Solving for these derivatives,

$$\frac{\partial g_z}{\partial p} = \frac{a_c + b}{a_c}\frac{\alpha_{ct} - 1}{\alpha_{ct}k(\tau_{ct})}exp((\alpha_{ct} - 1)p_{ct} + 1) > 0$$

$$\frac{\partial g_\mu}{\partial d} = \frac{\chi + exp(\alpha_{ct}p_{ct})}{k(\tau_{ct})} > 0$$

$$\frac{\partial g_\mu}{\partial p} = \frac{d_{ct}}{k(\tau_{ct})}\alpha_{ct}exp(\alpha_{ct}p_{ct}) > 0$$

$$\frac{\partial g_\varepsilon}{\partial d} = -(\chi + exp(\alpha_{ct}p_{ct}))\left\{exp(-\alpha_{ct}p_{ct}) + P_{ct}\theta_c\left[1 - exp(-\alpha_{ct}p_{ct})\right]\right\} < 0$$

Finally,

$$\frac{\partial g_\varepsilon}{\partial p} = -d_{ct}\alpha_{ct}\left[P_{ct}\theta_c\left[exp(\alpha_{ct}p_{ct}) + \chi exp(-\alpha_{ct}p_{ct})\right] - \chi exp(-\alpha_{ct}p_{ct})\right]$$

Notice that under no Prohibition, $\frac{\partial g_\varepsilon}{\partial p} > 0$ for any value of p_{ct}. Under Prohibition, a sufficient condition for $\frac{\partial g_\varepsilon}{\partial p} < 0$ (so that total crime is increasing in law enforcement) is that $\theta_c > \frac{\chi}{\chi + e_{ct}^{2\alpha}p_{ct}}$. In this case, $g(p_{ct}, d_{ct}, q_{ct})$ is one-to-one, and $|J|$ reduces to $|J| = \frac{\partial g_z}{\partial p}\frac{\partial g_\mu}{\partial d}\frac{\partial g_\varepsilon}{\partial q} = |J| = \frac{\partial g_z}{\partial p}\frac{\partial g_\mu}{\partial d}$. Thus, the likelihood function takes the form

$$\mathcal{L}_{ct}(\boldsymbol{y}_{ct}; \Theta_S, \theta_c, B_c, a_c, b, \alpha_{ct}, \chi, \overline{z}_{ct}, k, \lambda, \sigma_q^2, \sigma_z^2 | \varrho^{med}(P_{ct}), P_{ct}, \tau_{ct}) =$$

$$\frac{[g_\mu(\boldsymbol{y}_{ct})]^{a_c-1}[1 - g_\mu(\boldsymbol{y}_{ct})]^{b-1}}{\int x^{a_c-1}(1-x)^{b-1}dx}\frac{1}{2\pi\sigma_q\sigma_z}exp(-\frac{1}{2\sigma_q^2}g_\varepsilon(\boldsymbol{y}_{ct})^2)exp(-\frac{1}{2\sigma_z^2}(g_z(\boldsymbol{y}_{ct}; \varrho^{med}(P_{ct})) - \overline{z}_{ct})^2)\frac{\partial g_\mu(\boldsymbol{y}_{ct})}{\partial d}\frac{\partial g_z(\boldsymbol{y}_{ct})}{\partial p} \quad (25)$$