Estimating the Benefits of Policy Decentralization*

Andrew Arnold

University of Pennsylvania

Holger Sieg

University of Pennsylvania and NBER

July 21, 2022

*We would like to thank Steve Coate, Dennis Epple, Matias Iaryczower, Adam Meirowitz, Andrew Postlewaite, Kristopher Ramsay, Koleman Strumpf, and participants at numerous conferences and seminars for comments and suggestions.

Abstract

The objective of this paper is to develop and implement a new framework for estimating the benefits of policy decentralization. To accomplish this task, we consider a probabilistic voting model. We show how to identify and estimate this model using a combination of roll call data from state legislators and local votes in policy referenda. We study the most famous natural experiment ever conducted with respect to policy decentralization in the U.S., namely the decentralization of liquor policies at the end of the Prohibition Era. Compared to the Prohibition Era status quo, we find that aggregate welfare increases by 1 percent under the optimal uniform centralized policy. Decentralized policies offer opportunities to account for heterogeneity in preferences and increase welfare even further. Compared to the optimal uniform centralized policy, the optimal decentralized policy increases aggregate welfare by 36 percent.

KEYWORDS: Decentralization, Probabilistic Voting, Public Referenda, Roll Call Votes, Welfare Analysis, Prohibition Era.

1 Introduction

The temperance movement in the U.S. was dedicated to promoting moderation and, more often, complete abstinence from intoxicating liquor. The organized efforts supporting this movement involved religious coalitions that linked alcohol to virtually all of society's ills, including immorality, criminality, and lack of patriotism. The crowning achievement of the temperance movement was the passage of the 18th Amendment to the United States Constitution in 1917.[1] This amendment prohibited the production, sale, and transport of "intoxicating liquors," but it did not provide an enforcement strategy. The National Prohibition Act, known informally as the Volstead Act, was enacted by the United States Congress in 1919 to establish and enforce a uniform liquor control policy throughout the U.S.

Neither the Volstead Act nor the 18th Amendment were enforced with great success. Rather, entire illegal economies such as bootlegging, speakeasies, and distilling operations flourished. The public appetite for alcohol remained strong, and, despite initial success in suppressing alcohol consumption in the early 1920s, alcohol consumption rebounded to approximately two-thirds of the pre-Prohibition level by the end of the 1920s (Warburton, 1932). Organized crime took over the production and distribution of liquor, which led to sharp increases in violent crime. Law enforcement efforts were soon considered insufficient to deal with organized crime cartels, changing beliefs about enforcement costs and attitudes towards liquor control policies (Garcia-Jimeno, 2016).

Responding to widespread disenchantment with Prohibition Era policies, the U.S. Congress passed the 21st Amendment in 1933 voiding the Volstead Act. The political compromise that ended the Prohibition Era specified that liquor policies were no longer decided at the federal level. Instead, each state could determine its own policy. This com-

[1] The 18th Amendment was ratified and became effective in 1919.

promise resulted in the most famous natural social experiment ever conducted concerning policy decentralization in the U.S. Approximately half of the states further delegated the decision-making power to county governments or local municipalities. States with more heterogeneous preferences and strong minorities were more likely to decentralize liquor policies after the end of the Prohibition Era (Strumpf and Oberholzer-Gee, 2002).

Texas was the largest state in the Union that decided to decentralize liquor policies. Almost all relevant liquor policy votes were taken in the Texas legislature between 1931 and 1937. During that period we have identified 102 roll calls on bills and amendments that dealt with alcohol and liquor policies. Using standard scaling techniques, we can estimate the legislators' ideal points based on the observed roll call votes.[2]

We can link legislators' bliss points to voters' unobserved preferences using a probabilistic voting model.[3] We treat liquor policy as a "pliable issue" on which legislators are free to take any position they want in order to appeal to their constituents.[4] We show that the probabilistic voting model is identified up to some normalizations and can be estimated using a GMM estimator. Our preferred model defines voter types based on religious affiliations, classifying voters into "wet" and "dry" types.[5] Using data on

[2]See, for example, Poole and Rosenthal (1985, 1991), Heckman and Snyder (1997), or Clinton, Jackman, and Rivers (2004). While such measures are subject to some criticism, they have proven to be useful in many applications. Moreover, different estimation techniques often result in similar estimates of the ideal points. Our empirical framework method can be combined with any method that provides reasonable measures of the legislators' ideal points.

[3]For a comprehensive survey and analysis of probabilistic voting models see Coughlin (1992).

[4]We show below that there was no party discipline within the Texas Democratic party, which dominated state politics during the relevant period we study. In the language of probabilistic voting models, liquor policy was not a fixed issue which was dominated by the party.

[5]Garcia-Jimeno, Iglesias and Yildirim (2021) study the role of information networks and collection action during the Temperance Crusade in 1873-74. They also highlight the importance of religious organizations such as Women's Christian Temperance Union in the push for prohibition.

the socio-economic and demographic compositions of voting districts, as well as the estimated bliss points of each member of the Texas House of Representatives, we estimate the parameters of the probabilistic voting model. We find that the estimates are quite reasonable and that the model fits the data well. We validate the model using the estimated bliss points of members of the Texas Senate which were not used in the estimation of the probabilistic voting model.

Liquor policies in Texas were ultimately determined by local public referenda, which pit a status quo against an alternative policy (Romer and Rosenthal, 1978). Given that we have identified voters' preferences from the probabilistic voting model of representation, we can also identify the unobserved policy positions of the status quo and the proposed alternative from the observed vote shares of the public referenda. We observe the outcome of 302 referenda at the county level between 1937 and 1952.[6] We find that our model explains the observed vote shares and provides reasonable estimates for the policies that were subject to the referenda.

Once we have estimated and validated our model, we can compare voters' welfare under different policy regimes. First, we compare the dry status quo of the Prohibition Era to the optimal uniform policy. We find that aggregate welfare increases by only 1 percent under the optimal uniform centralized policy. We thus conclude that the Prohibition Era dry policy was fairly close to the optimal uniform policy. We then compare the Prohibition Era status quo and the optimal decentralized policy. We find that aggregate welfare increases by 36 percent under the optimal decentralized policy. Hence, there are large and significant benefits from policy decentralization. Finally, we

[6]These data were also used by Strumpf and Oberholzer-Gee (2002) and Coate and Conlin (2004). See these papers for a more detailed discussion of these data. The vast majority of referenda pitted the dry Prohibition Era status quo against a "beer & wine only" alternative. We observe few referenda that allowed the sale of all types of liquor in Texas.

document that there existed much heterogeneity in preferences across voting districts and counties in Texas and conduct a disaggregate analysis of gains and losses under various policy regimes. Our empirical results thus provide a clean interpretation of the failures of the Prohibition Era and provide plausible estimates of the magnitudes of the welfare gains that can arise from decentralization.

Our paper is related to various parts of the literature on political economy and fiscal decentralization. One of the main insights of fiscal federalism is that a decentralized provision of public goods is likely to improve welfare over a centralized, uniform provision if spillovers and differences in costs among local jurisdictions are negligible (Oates, 1972). Given the prevalence of policy decentralization in most developed economies, it is surprising that there are few compelling empirical studies that have estimated the benefits of policy decentralization within an internally consistent political economy model. One exception is Calabrese, Epple, and Romano (2011) who compare the efficient decentralized allocation, the equilibrium allocation with decentralized provision, and the equilibrium allocation with a uniform provision in a model that is calibrated to the Boston metropolitan area. This paper develops and implements a completely different framework for measuring the benefits of policy decentralization. One important methodological contribution is that we show how to estimate the parameters of a probabilistic voting model and characterize the position of various policies that were implemented by combining data that characterize votes taken by legislators and vote shares in local policy referenda.

Another seminal empirical paper on fiscal federalism and policy decentralization is Strumpf and Oberholzer-Gee (2002). They study whether states with high preference heterogeneity and strong minorities were more likely to decentralize liquor policies after the end of the Prohibition Era. Our research differs from this paper in several important ways. First, Strumpf and Oberholzer-Gee (2002) estimate preferences of legislators by exploiting variation in decentralization decisions across states. In contrast, we exploit variation in the preferences of legislators within a single state. Second, we rely on an explicit model of political competition that links voters' preferences to legislators' preferences while Strumpf and Oberholzer-Gee (2002) derive the key estimation equations from a model of legislative decision-making. We also test whether voters' behavior in referenda is consistent with the outcomes of legislative elections. We validate the model using data from senators. Finally, our analysis allows us to ultimately compare welfare differences between centralized and decentralized decision-making, whereas Strumpf and Oberholzer-Gee (2002) are primarily interested in establishing that heterogeneity within states predicts policy decentralization.

The decentralization theorem has been criticized since it assumes that centralization results in a uniform policy. In general, federal governments can provide different levels of public services among states. Nevertheless, there are often political and constitutional constraints on the extent to which this can happen. The theoretical literature has developed models of decentralization that have endogenized the centralized policy. Lockwood (2002) considers a model of legislative bargaining and shows that more heterogeneity among preferences across jurisdictions does not necessarily imply that welfare increases under decentralization. Besley and Coate (2003) have shown that decentralization is desirable under the same conditions if one relaxes the assumption that the federal government is constrained to implement a uniform policy. In their model centralization tends to generate an over-provision of public goods because voters have incentives to elect representatives with high demand for public spending. In these models, voters do not necessarily want to elect candidates who share their policy preferences. Depending on the composition and procedures of the legislature, they may want to strategically delegate to candidates with more extreme preferences to counteract the influence of other districts. Our analysis abstracts from these strategic voting issues.

These papers are also motivated by the observation that federal government policies are not necessarily uniform. In contrast, we study an era in which it was the explicit objective of the federal government to implement a uniform policy for all states. As discussed above, both the 18th Amendment to the United States Constitution and the National Prohibition Act had the intention to establish and enforce a uniform liquor policy throughout the U.S. It thus makes sense to compare decentralized policies against uniform centralized policies in our application.

Our paper is also related to the literature that has studied the validity of voting models using state and local data. Compelling indirect evidence in favor of the median voter theorem is provided by Lott and Kenny (1999) who examine the growth of state governments as a result of giving women the right to vote. Similarly, Miller (2008) finds that suffrage rights for American women helped children to benefit from technological innovations in health care and significantly decreased child mortality in the U.S. Epple, Romer, and Sieg (2001) and Calabrese, Epple, Romer, and Sieg (2006) analyze whether observed local tax and expenditure policies are consistent with a version of the median voter theorem that accounts for endogenous household sorting within a system of jurisdictions. In contrast to these papers, we use a probabilistic voting model and directly exploit voting data to identify, estimate, and validate our model.

Our paper is also related to Coate and Conlin (2004) who estimate models of voter turnout also using data from Texas liquor referenda. They find that a rule-utilitarian model provides a good explanation for the observed turnout patterns. This result is important since it provides a compelling explanation of why voters show up at the ballot box. Coate, Conlin, and Moro (2008) extend the analysis and estimate a pivotal voting model. We do not provide an explicit model of voter turnout. However, we show how our model and empirical analysis can also account for imperfect voter turnout.

Finally, our paper adds to the recent literature on estimating game-theoretic models in political economy. An early example of theory-based estimation in political economy is Merlo (1997), who estimates a dynamic bargaining model of government formation. More recently, Knight and Schiff (2010) estimate a model of social learning in presidential primaries. Iaryczower and Shum (2012) estimate a game with asymmetric information to describe the voting behavior of judges in appellate courts. Kawai and Watanabe (2013) consider models of strategic voting. Sieg and Yoon (2017, 2022) estimate dynamic games with asymmetric information of electoral competition and evaluate the impact of term limits. Aruoba, Drazen, and Vlaicu (2019) use governors' job approval ratings as outcome measures and estimate a model with moral hazard. Battaglini, Patacchini, and Rainone (2020) estimate a model of network formation and show that social connections are important for legislators' productivity. Our model and estimation strategy significantly differs from all those papers.

The rest of the paper is organized as follows. Section 2 discusses the institutional background and introduces our data set. Section 3 develops the theoretical framework for our analysis. Section 4 discusses the identification and estimation of our model. Section 5 presents the main empirical results. Section 6 explores the policy and welfare implications of our work. Section 7 concludes.

2 Data

2.1 Historical and Institutional Background

Since 1876, Texas has allowed localities to determine their liquor statuses through local option elections and referenda. As discussed above, the 18th Amendment to the U.S. Constitution on National Prohibition was ratified in January 1919. Texans followed suit with a Prohibition amendment to the state constitution in May 1919.

Overall, the impact of Prohibition on alcohol consumption in the U.S. is difficult to ascertain. While the federal government collected reliable data on alcohol consumption before and after the Prohibition Era, no accurate government data are available for the Prohibition Era. Perhaps more surprisingly, Warburton (1932) is the only careful economic study that was ever conducted towards the end of the Prohibition Era to determine the impact of Prohibition on alcohol consumption and a variety of other economic outcomes.

Figure 1: U.S. Alcohol Consumption: 1900-1955

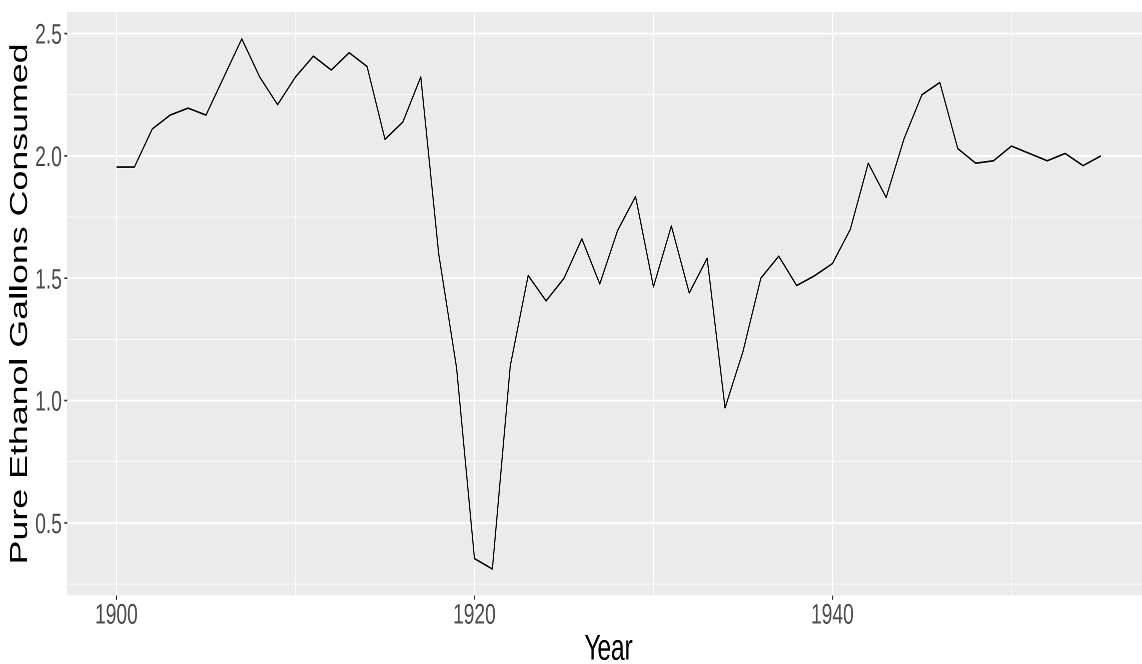


Figure 1 shows the time series of total pure ethanol consumption per capita (using the population 15 and older as deflator) between 1900 and 1955. Our data are based on the official statistics for the periods 1900-1920 and 1934-1955. For the years 1921-1930 we rely on the estimates from Warburton (1932, Table 30).[7] Note that alcohol consumption was fairly stable before the Prohibition Era and averaged around 2.39 gallons per capita

[7]These estimates were confirmed by Miron and Zwiebel (1991) who extended Warburton's analysis

during the period between 1911-14.[8] Using a variety of different techniques, Warburton (1932) estimated that alcohol consumption fell sharply at the beginning of the Prohibition Era, to approximately 25-30 percent of its pre-Prohibition level. By the mid-1920s, however, law enforcement efforts slackened, leading to a significant rebound in alcohol consumption that reached approximately 60-70 percent of its pre-Prohibition level in 1930. The level of consumption remained the same right after the end of the Prohibition Era and averaged about 75-80 percent of its pre-Prohibition level between 1935 and 1955.

While the aggregate time series for pure ethanol illustrates the main trend in alcohol consumption, it conceals some important aspects of the Prohibition Era. Table 7 in Appendix D reports three time series breaking down the total pure ethanol consumption into beer, wine, and spirits consumption. Notably, Prohibition primarily succeeded in reducing beer consumption. The per capita consumption of beer was reduced by about 70 percent, from approximately 31 gallons of beer per year (or 1.25 gallons of pure ethanol) in 1911-1914 to 8.75 gallons of beer (or 0.35 gallons of pure ethanol) in 1927-1930. In contrast, consumption of wine increased by 65 percent and pure spirits by 10 percent during the Prohibition Era. As a consequence, Prohibition led to a strong substitution from beer to wine and spirits. This was probably due largely to the fact that it was much easier to enforce Prohibition for large-scale beer production by shutting down commercial breweries and forcing households into illegal home production of beer.[9]

using data from the post-Prohibition Era and conducting a variety of sensitivity checks. Overall, their findings are essentially the same as those reported in Warburton (1932). Miron and Zwiebel (1991) also provide estimates for the period from 1931-33 which are used in the graph above.

[8] Warburton (1932) reports on p. 147 that the federal tax rates on beer were increased in 1917 from \$1.50 per barrel to \$3, and in 1919 to \$6, and those on spirits from \$1.10 per gallon to \$3.20 in 1917, and to \$6.40 in 1919. The rate on wines, which previously had borne no federal tax, was made to vary with the alcoholic content. This partially explains the large drop in alcohol consumption during World War I.

[9] There are no estimates of alcohol consumption at the state level for the Prohibition Era. However,

Prohibition remained in effect for almost 14 years until it was rescinded by the 21st Amendment in 1933. In 1933, Texans also voted to repeal Prohibition and voted in favor of allowing the sale of 3.2 percent beer, but these changes had no immediate effect because of the still-active statewide Prohibition law. The state dry law was repealed in August 1935, returning each county to its pre-Prohibition liquor status. The statewide sale of alcoholic beverages was licensed in 1936. The first post-Prohibition local option elections were recorded in 1937 when over 70 localities voted to change their alcohol status. By 1953, local liquor statuses were mostly constant, with an average of only 17 elections per year between 1953 and 1958. We, therefore, focus on public referenda between 1933 and 1952.

Texas uses a bicameral system which consists of a Senate and a House of Representatives. During the relevant period, there were 31 single-member Senate districts and 150 House districts. Each senatorial district consisted of a collection of counties.[10] House districts and Senate districts did not cut county lines. Unlike House districts, no county appeared in more than one Senate district, i.e. each senator was elected from a single-member district. In contrast, representatives were elected from single- and multi-member districts.[11] Between 1933 and 1952 the Texas House of Representatives had 150 members from 127 districts. In particular, 115 representatives came from single-member districts

more recent statistics suggest that Texas tends to be in the 5th or 6th decile of alcohol consumption among the 50 states in the U.S. As a consequence, the US data also provide an accurate estimate of alcohol consumption in Texas.

[10]Texas is divided into 254 counties.

[11]Multi-member districts were used extensively from 1846 until 1975. Although the idea of equal representation surely held some conceptual appeal, it was not until 1964 that the Supreme Court insisted on the “one person, one vote” principle for state legislative districts in the landmark Reynolds v. Sims decision (Calabrese, 2000). This and other court decisions, as well as the difficulty of developing a district plan that would pass muster, led to the exclusive use of single-member districts in Texas after 1975 (Eckham, 2016).

accounting for 72.3 percent of all representatives. The remaining 35 representatives came from 8 two-member districts, one four-member district, and three five-member districts. The districts usually corresponded to counties or combinations thereof, with more populous counties getting more representatives.[12] Only two of the Representatives or Senators in office from 1933 to 1937 were unaffiliated with a party, and one of them later became a Democrat. All others were Democrats. Hence, the main political competition was within the Democratic party.

2.2 Data Sources and Descriptive Statistics

Using the 1940 U.S. Census micro-data, we observe several socio-economic demographics characterizing the income, age, gender, and racial distributions of the districts.[13] Moreover, we observe variables characterizing labor and housing markets such as wage income, fraction unemployed, housing wealth, and fraction renter. We also have detailed information about the different religious affiliations at the county level from the 1936 Decennial Census of Religious Bodies. We follow Garcia-Jimeno (2016) in grouping these religious affiliations into two types: "dry" and "wet". This grouping reflects the prevailing attitude of these religious groups towards liquor policies. Dry religions are Evangelicals, Baptists, Methodists, Presbyterians, and Mormons. Wet religions are Lutheran, Catholic, and Jewish.[14] We also observe the total population and land area which allows us to construct a population density measure. Table 1 provides socio-demographic characteristics at the district level for both members of the House of Representatives and the Senate.

[12]From time to time, flotorial districts were used. These were districts that overlapped with other districts to give representation to a county or region that otherwise would not get any.

[13]We use the Integrated Public Use Micro-data Series version of the complete count Census data. The 1936 Decennial Census of Religious Bodies comes from ICPSR Study #8.

[14]Garcia-Jimeno (2016) also includes Eastern Orthodox as a wet religion, but there is no Eastern Orthodox community in Texas in the 1936 Census.

Table 1: House and Senate Districts

	House		Senate	
	Mean	Std Dev	Mean	Std Dev
Demographics				
Population ($\times 10^5$)	0.95	1.00	5.30	0.19
% Male	50.45	0.92	50.29	0.79
% Black	15.70	13.27	15.60	11.34
% Hispanic	7.96	15.00	7.65	12.95
% Black or Hispanic	23.60	16.37	23.19	13.60
% White	76.38	16.38	76.78	13.61
% 18+	64.74	3.23	65.65	3.13
% 21+	58.85	3.36	59.76	3.25
Average wage income ($\times 10^3$ 1940 \$)	0.70	0.16	0.71	0.15
Median home value ($\times 10^3$ 1940 \$)	0.89	0.49	0.97	0.56
Population Density ($\times 10^3$) per square mile	0.45	0.52	0.55	0.78
% Renter	57.74	5.49	57.57	3.80
% Urban	32.40	21.68	38.26	20.20
% Unemployed	7.23	2.19	7.34	1.61
% Completed high school	24.76	7.68	26.97	7.47
% Some college	10.62	3.32	11.54	2.92
Dry Religions				
% Evangelical	0.68	1.81	0.75	1.35
% Baptist	39.53	17.06	38.03	13.82
% Methodist / Episcopal	21.46	8.96	21.17	7.86
% Presbyterian	3.59	2.22	3.70	1.64
% Mormon	0.00	0.00	0.00	0.00
Wet Religions				
% Lutheran	3.52	6.93	3.59	4.83
% Jewish	0.64	1.46	1.18	2.07
% Catholic	18.26	22.05	19.26	20.62

Table 2: Roll Call Votes on Liquor Policy

Chamber	Number of Legislators	Number of Single Member Districts	Number of Roll Calls Liquor Policy	Mean Liquor Votes
House	226	173	102	35.82
Senate	31	31	93	47.90

We measure the ideal points of legislators using the well-known scaling techniques developed by Poole and Rosenthal (1985, 1991). For each legislator in the House and Senate, we only analyze the votes on liquor policies.[15]

We can identify roll call votes on this policy dimension using the journals of the House and the Senate made available by the Texas Legislative Reference Library. These journals index all proposed legislation by their policy content and contain detailed histories of each bill, including an overview of every time the legislature amends or votes to pass or table the bill. The literature suggests that twenty is the minimum number of votes necessary for obtaining reliable estimates of the bliss points.[16] Table 2 shows that there were 226 members of the House that voted at least 20 times on liquor policies during the relevant period. Moreover, 173 of these representatives came from single-member districts. Similarly, we have identified 31 senators that voted at least 20 times on liquor

[15] Since we only use roll call votes on liquor policies, we estimate a one-dimension model. As we will discuss below this procedure is justified if preferences are additively separable across policy dimensions. It is straightforward to extend our approach to a multi-dimensional policy space. We have also analyzed voting on a fiscal dimension and the results are available upon request from the authors. If we made additional assumptions and label the votes as ”wet” and ”dry”, an alternative is to estimate a linear probability model of ”vote wet” on member and vote fixed effects.

[16] See, for example, McCarty, Poole, and Rosenthal (2006).

policy. This sample contains the vast majority of all Representatives and Senators that served during the relevant period in the Texas legislature.

Figure 2: Histogram of Alcohol Bliss Points

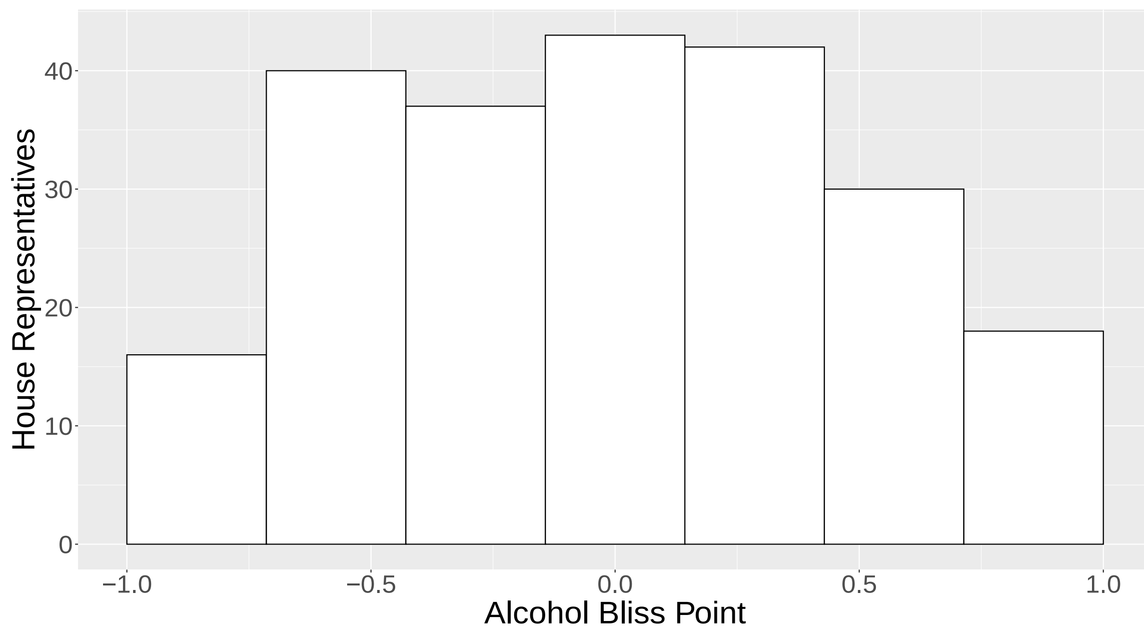


Figure 2 provides a histogram of the Nominate estimates of the liquor policy ideal points for the members of the Texas House of Representatives in our sample. We find that there was much heterogeneity in preferences over liquor policy in the House of Representatives. We also estimated the bliss points for 31 senators that are used as a hold-out sample as part of a validation exercise. Overall, we find that Senators had more extreme liquor policy bliss points than members of the House of Representatives. The distribution of the bliss point strongly suggests that legislative votes on liquor policies were not subject to party discipline. Instead, each legislator had the freedom to vote in accordance to his own preferences and those of his constituencies.

Liquor referenda were either held at the county, justice precinct, city, or town level.

Since voting districts are composed of counties in Texas, we focus on the 302 county-level referenda during our period. We observe 122 unique counties holding local option elections between 1937 and 1952. These 122 counties comprised over 48 percent of the total number of counties in Texas at the time. All referenda can be broadly classified into three types: a) beer-only voted out (in), b) 14 percent beverages voted out (in); and c) all alcoholic beverages voted out (in). We observe 183 beer-only versus dry referenda, 38 beer-and-wine-only versus dry referenda, and 81 dry versus all referenda. For counties that held multiple referenda, we computed an average vote share. Table 3 provides some descriptive statistics for our sample.

Table 3: Local Referenda: Descriptive Statistics

	Individual Referenda		Average Referenda	
	Mean	Std Dev	Mean	Std Dev
Vote Share	42.14	12.85	41.34	14.20
Turnout	27.56	13.00	26.47	11.04
Number of Referenda	302		122	
% of Counties that Held a Referendum	48.03		48.03	

The mean share was 42 percent, indicating that the majority of referenda were in favor of keeping the county dry. The mean voter turnout was approximately 27 percent.

3 A Probabilistic Voting Model

3.1 Voting in Two-Candidate Elections

The starting point of our analysis is a model of political competition and representation at the district level. We adopt a standard probabilistic voting model with two candidates that compete to represent the district in the legislature.[17] Let us assume that there are I types of voters. Each type has a share given by s_i. A policy is given by a K-dimensional vector, $g = (g_1, ..., g_K)$. For empirical tractability, we assume that preferences of voter i are additively separable over the K policy dimensions and given by:

$$u_i(g) \;=\; \sum_{k=1}^{K} u_{ki}(g_k) \;=\; \sum_{k=1}^{K} - \,\omega_{ki}\,(g_k - \theta_{ki})^2 \tag{1}$$

Hence, θ_{ki} is the bliss points of voter group i for the policy dimension k, and ω_{ki} is the relative weight that voter i assigns to the policy dimension k.

There are two candidates a and b that are competing for office. Each candidate can fully commit to a policy position before the election. Voters' utilities are subject to random shocks, denoted by ϵ_{ij}, which are not observed by the two candidates. Hence, the share of voters of type i that vote for candidate a is given by:

$$\begin{aligned} V_i^a(g_a, g_b) \;&=\; Pr\{u_i(g_a) + \epsilon_{ia} \geq u_i(g_b) + \epsilon_{ib}\} \\ &=\; \Phi_i(u_i(g_b) - u_i(g_a)) \end{aligned} \tag{2}$$

where $\Phi_i(\cdot)$ is the distribution function of $\epsilon_{ib} - \epsilon_{ia}$. The total vote share of candidate a is, therefore, given by:

$$V^a(g_a, g_b) \;=\; \sum_i \; s_i \; \Phi_i(u_i(g_a) - u_i(g_b)) \tag{3}$$

and $V^b(g_a, g_b) = 1 - V^a(g_a, g_b)$.

Each candidate is maximizing her vote share. Taking derivatives with respect to g_{ka}, we obtain for each k:

$$\frac{\partial V^a(g_a, g_b)}{\partial g_{ka}} \;=\; -2 \sum_i \; s_i \; \phi_i(u_i(g_a) - u_i(g_b))\,\omega_{ki}\,(g_{ka} - \theta_{ki}) \;=\; 0 \tag{4}$$

[17] Coughlin (1992) provides a detailed discussion of probabilistic voting theory. One of the drawbacks of probabilistic voting theory is that it assumes commitment. Citizen candidate models that are due to Osborne and Slivinski (1996) and Besley and Coate (1997) provide one way to relax the commitment assumption.

and a similar condition for g_{kb}.

In a symmetric Nash equilibrium, both candidates commit to pursue the same policies, implying that $u_i(g_a) - u_i(g_b) = 0$.[18] Hence, we obtain:

$$\sum_i s_i \; \phi_i(0) \; \omega_{ki} \; \theta_{ki} \quad = \quad g_k \sum_i s_i \; \phi_i(0) \; \omega_{ki} \tag{5}$$

Solving this equation implies that the optimal position of each candidate for policy k is given by:

$$g_k \quad = \quad \sum_i \frac{s_i \; \phi_i(0) \; \omega_{ki}}{\sum_j \; s_j \; \phi_j(0) \; \omega_{kj}} \; \theta_{ki} \tag{6}$$

Once in office, a legislator has quadratic, additively-separable preferences with bliss point given by g_k. Each legislator votes sincerely on each policy proposal. Note that the bliss point is a weighted average of the bliss points of the voters for each policy dimension. The weights depend on the share of each type (s_i), the density of the median voters ($\phi_i(0)$), and the relative importance of the policy dimension (ω_{ki}).

The probabilistic voting model predicts that both candidates take the same position on all pliable issues on which legislators are free to take any position they want in order to appeal to their constituents. In our application, we have seen that the Democratic party dominated state politics. We should, therefore, interpret the model above as a theory of political competition within a party. We do not find any evidence that the Democratic party exercised strong party discipline on liquor policy during the post-Prohibition Era that we study in this paper.[19] As such we conclude that the model provides a a reasonable mapping of voters' preferences into legislators' bliss points. Since we can estimate legislators' bliss points, we primarily use the model to invert this mapping and recover the unobserved preferences of voters.

[18]With a slight abuse of notation let use denote this equilibrium policy by g.

[19]Since we do not observe candidates from different parties, we cannot conduct the standard test whether legislators with different party affiliations from the same district pursue different policies.

3.2 Voting in Local Referenda

Next, we consider voting in local public referenda. We assume that each proposal only affects a single dimension k. A referendum pits a status quo, denoted by g_{ks}, against an alternative policy, denoted by g_{kr}. Since preferences are additively separable, voter i prefers the alternative policy if and only if:

$$-(g_{kr} - \theta_{ki})^2 \; + \; \epsilon_{kri} \geq -(g_{ks} - \theta_{ki})^2 \; + \; \epsilon_{ksi} \tag{7}$$

where ϵ_{ksi} and ϵ_{kri} are random election shocks that impact the vote during the referendum. Hence, the share of votes in favor of the alternative policy proposed in the referendum is given by:

$$V(g_{ks}, g_{kr}) \;\; = \;\; \sum_i \; s_i \; F_{ki}((g_{ks} - \theta_{ki})^2 - (g_{kr} - \theta_{ki})^2) \tag{8}$$

where F_{ki} is the distribution function of the referendum shock $\epsilon_{ki} = \epsilon_{ksi} - \epsilon_{kri}$. Note that vote shares in referenda do not depend on ω_{ki}

3.3 Centralization versus Decentralization

Finally, we discuss how to construct welfare measures that allow us to compare the outcomes under centralization and decentralization. Consider a model with voting districts $n = 1, ..., N$. The voter's utility function in equation (1) is additively separable across policy dimensions. As a consequence, we can construct welfare measures that are also additively separable across policy dimensions. Recall that voter i in district n has preferences over the kth policy dimension given by

$$u_{kin}(g_k) \;\; = \;\; -\omega_{ki} \; (g_k - \theta_{kin})^2 \tag{9}$$

The status quo is denoted by g_{ks}. Assuming a utilitarian welfare function, a welfare measure of district n under the status quo is then

$$W_{ns}(g_{ks}) \;=\; \sum_{i=1}^{I} s_{in}\; u_{kin}(g_{ks}) \tag{10}$$

where s_{in} is district n's share of voter type i. Total welfare for all N voting districts is then given by

$$W_{ks} \;=\; \sum_{n=1}^{N} w_n\; W_{ns}(g_{ks}) \tag{11}$$

where w_n is the population share of district n.

The optimal uniform centralized policy, g_{kc}, that maximizes aggregate welfare is:

$$g_{kc} \;=\; \text{argmax}_{g_k \in [-1,1]} \;\sum_{n=1}^{N}\; w_n \;\sum_{i=1}^{I} s_{in}\; u_{kin}(g_k) \tag{12}$$

It is straightforward to show that g_{kc} is the population-weighted average of the districts' voter share-weighted alcohol bliss points. Total welfare for all N voting districts is then given by

$$W_{kc} \;=\; \sum_{n} w_n\; W_{ns}(g_{kc}) \tag{13}$$

The optimal decentralized policy, g_{kdn}, maximizes welfare for each district n, $W_n(g_k)$:

$$g_{kdn} \;=\; \text{argmax}_{g_k \in [-1,1]} \;\sum_{i=1}^{I}\; s_i\; u_{kin}(g_k) \tag{14}$$

It is straightforward to show that it is given by the district's voter share-weighted bliss point

$$g_{kdn} \;=\; \sum_{i} s_{in}\; \omega_{ki}\; \theta_{kin} \tag{15}$$

Aggregate welfare under the optimal decentralized policy is then

$$W_{kd} \;=\; \sum_{n=1}^{N} w_n\; W_{dn}(g_{kdn}) \tag{16}$$

In summary, we have shown how to characterize the welfare of alternative policies. We have discussed how to compute the optimal policies under centralization and decentralization. One main empirical objective of this analysis is to implement these welfare measures. To accomplish this task, we need to identify and estimate the parameters of the probabilistic model. To compute changes in welfare relative to the Prohibition Era status quo, we also need to estimate the position of policies that have been implemented in the past. We discuss these issues in the next sections of the paper.

4 Identification and Estimation

The intuition for the empirical strategy proposed in this paper is fairly straightforward. Recall that we can estimate the legislators' bliss points using standard scaling techniques as long as we observe a sufficiently large number of roll call votes. We have discussed in Section 3.1. how to endogenize the bliss point of a legislator who represents a voting district. In particular, equation (6) provides the crucial link between the (estimated) bliss point of the legislator and the (unobserved) preferences of the voters in the district. We also observe the vote shares in local liquor referenda. In Section 3.2, we characterized how citizens vote in referenda that pit a status quo against an alternative proposal. In particular, equation (8) provides the link between the (observed) vote shares and the (unobserved) positions of the status quo and the alternative policy. That allows us to characterize the policies that were under consideration and implemented in practice and compute the welfare associated with these policies. Once we have identified and estimated the preferences of the voters in each district, we can also estimate the optimal uniform policy and the vector of optimal decentralized policies. We discuss the details of our identification and estimation strategy in detail below.

Consider a sample that consists of a large number of districts, $n = 1, ..., N$. We can classify voters into types based on observed characteristics such as age, race, and gender. Hence, we observe the shares s_{in} for $i = 1, ..., I$. We consider an environment with a small I and a large N.

We use a sequential approach to identification. Moreover, our proofs of identification are constructive and give rise to a GMM estimator. Recall that we can identify the bliss points of the legislators that represent each district, denoted by g_{kn}, based on roll call votes using standard scaling techniques. Here, we show that we can identify the structural parameters of the legislative voting model based on equation (6). We then show how to identify the remaining parameters of the model using vote shares for liquor referenda and exploiting equation (8).

Let us assume that we observe some characteristics of the district, denoted by x_{kn}, that systematically shift preferences over policies of each group. These are socio-economic characteristics that are not used to define voter types. Note that we allow these characteristics to differ by the dimension of the policy space. Let us assume that the bliss points of each voter type satisfy:

$$\theta_{kin} \; = \; \alpha_{ki} + \beta_{ki}\, x_{kn} \; + \; u_{kin} \tag{17}$$

where u_{kin} is an error term. Substituting equation (17) into equation (6), we obtain:

$$g_{kn} \; = \; \left(\sum_i \frac{s_{in}\, \gamma_{ki}}{\sum_j \; s_{jn}\, \gamma_{kj}} \, (\alpha_{ki} + \beta_{ki}\, x_{kn}) \right) \; + \; v_{kn} \tag{18}$$

where $\gamma_{ki} = \phi_i(0)\, \omega_{ki}$, and where

$$v_{kn} \; = \; \sum_i \frac{s_{in}\, \gamma_{ki}}{\sum_j \; s_{jn}\, \gamma_{kj}} \, u_{kin} \tag{19}$$

Assuming that $E[u_{kin}|\; s_{in}, x_{kn}] = 0$ for all k, i, and n, we have

$$E[v_{kn}|x_{kn}, s_{1n}, ..., s_{In}] \; = \; \sum_i \frac{s_{in}\, \gamma_{ki}}{\sum_j \; s_{jn}\, \gamma_{kj}} \, E[u_{kin}|x_{kn}, s_{in}] \; = \; 0 \tag{20}$$

Assuming that $E[u_{kin}^2 | \, s_{in}, x_{kn}] = \sigma_{ki}^2$ for all k, i, we have

$$Var[v_{kn}|s_{1n}, ..., s_{In}] \;\; = \;\; \sum_i \left(\frac{s_{in}\gamma_{ki}}{\sum_j s_{jn}\gamma_{kj}} \right)^2 \sigma_{ki}^2 \tag{21}$$

We therefore have the following results:

Proposition 1 *The parameters* $(\alpha_{ki}, \beta_{ki}, \gamma_{ki})$ *are identified up to a normalization such as* $\sum_i \gamma_{ki} = 1$. *Moreover, the conditional distributions of* u_{kin} *are non-parametrically identified.*

The proof of Proposition 1 is given in Appendix A. We offer three observations. First, the proof is constructive and can be used to construct a GMM estimator that that is based on orthogonality conditions that can be constructed based on equations (20) and (21).[20] Second, recall that $\gamma_{ki} = \phi_i(0)\; \omega_{ki}$. Hence, we cannot separately identify $\phi_i(0)$ and ω_{ki}. Third, the normalization of the γ_{ki} just reflects the well-known fact that the absolute level of the utility can be normalized.

We also observe a large sample with size N of votes on the same type of referendum.[21] Adding subscripts and substituting equation (17) into the utility function, we obtain:

$$\begin{aligned} & (g_{ks} - \alpha_{ki} - \beta_{ki}x_{kn} - u_{kin})^2 \; - \; (g_{kr} - \alpha_{ki} - \beta_{ki}x_{kn} - u_{kin})^2 \\ = \;\; & g_{ks}^2 + 2\,(g_{kr} - g_{ks})\,(\alpha_{ki} + \beta_{ki}x_{kn}) \; + \; 2\,(g_{kr} - g_{ks})\,u_{kin} \; - \; g_{kr}^2 \end{aligned} \tag{22}$$

We assume that F_{ki} is Type I extreme value distributed. Hence, we can write the vote share equation as follows:

$$V_{kn} = E[V_{kn}|s_{1n}, ..., s_{In}, x_{kn}] + w_{kn} \tag{23}$$

[20]See Appendix B for a detailed derivation of the GMM estimator.

[21]We do not require that the sample size in the third stage is equal to the sample size in the first and second stages of the estimation. However, we use the same notation for expositional simplicity.

where $E[V_{kn}|s_{1n},...,s_{In},x_{kn}]$ is given by

$$\sum_i s_{in} \int \frac{\exp[(g_{ks}^2 + 2\,(g_{kr} - g_{ks})\,(\alpha_{ki} + \beta_{ki}x_{kn})\ +\ 2\,(g_{kr} - g_{ks})\,u_{ki}\ -\ g_{kr}^2)/\sigma_{ki}^{\epsilon}]}{1 + \exp[(g_{ks}^2 + 2\,(g_{kr} - g_{ks})\,(\alpha_{ki} + \beta_{ki}x_{kn})\ +\ 2\,(g_{kr} - g_{ks})\,u_{ki}\ -\ g_{kr}^2)/\sigma_{i}^{\epsilon}]}\, r(u_{ki})\, du_{ki}$$

and σ_i^{ϵ} is the scale parameter of the election shock of voter type i. Hence, the remaining parameters of the model, given by $g_{ks}, g_{kr}, \sigma_i^{\epsilon}$, are identified.[22] We can thus construct additional orthogonality conditions based on equations (23) and estimate the parameters of the voting model using a GMM estimator. Note that it is straightforward to extend the analysis and allow for more than one type of referenda, as long as we have a large enough sample size for each type.

For estimation purposes, it is useful to impose the constraint that θ_{kni} to be in the interval $[-1,1]$. We can achieve this by assuming that

$$\theta_{kni} = \left[\frac{2}{1 + \exp(-\alpha_{ki} - \beta_k x_{kn})} - 1\right] + u_{kni} \tag{24}$$

where u_{kni} is truncated normal.

In summary, the estimation proceeds in two steps. First, we estimate the bliss points of politicians in the legislature using standard spatial estimation techniques. Second, we estimate the parameters of the voters' preferences as well as the positions of the status quo and the alternative policy proposed in the referendum using a GMM estimator. We have conducted a Monte Carlo study and found that our estimator performs quite well using simulated data.[23]

[22]Up to a normalization that $\sigma_1^{\epsilon} = 1$.

[23]The Monte Carlo results are available upon request from the authors.

5 Empirical Results

5.1 Parameter Estimates

As we discussed in Section 2 of the paper, we classify voters into two types using religious affiliation. Type 1 voters are members of a “dry” religion, and type 2 voters are affiliated with a “wet” religion. Using this classification system, we then estimate a variety of different specifications of our model for the liquor policy dimension. The results are summarized in Table 4.[24] The specification in Column (1) only uses the orthogonality conditions that can be derived from the bliss points of the legislators, i.e. we do not use the orthogonality conditions based on the referenda. The sample consists of all representatives from single-member districts. The specification in Column (2) adds the orthogonality conditions based on the referenda. Since some counties had multiple referenda, we use the average wet vote share as the outcome variable. This is then our preferred model specification. We consider all referenda that pitted the dry status quo against an alternative. Column (3) includes bliss points from the 53 multimember district representatives. Column (4) excludes all referenda from 1942 onward. Column (5) excludes any referenda that did not pit “beer only” against the dry status quo.

Given the sequential nature of our estimation strategy, we use a bootstrap algorithm to estimate standard errors. We randomly drop one alcohol vote for each representative and reestimate the bliss points based on the leave-one-out sample. We then drop one representative-district observation and one referendum and reestimate the model on the leave-one-out samples. The estimated standard errors are then obtained based on 100 iterations of this jackknife algorithm.

[24] We also estimated a model for the fiscal dimension and the results are available from the authors.

Table 4: Parameter Estimates: Liquor Policy

	(1)	(2)	(3)	(4)	(5)
			Liquor Policy		
α_1 (Intercept Dry)	5.19	6.18	5.32	6.29	6.08
	(0.76)	(0.45)	(0.30)	(0.29)	(0.28)
α_2 (Intercept Wet)	1.80	0.83	1.72	0.74	0.96
	(0.76)	(0.36)	(0.22)	(0.25)	(0.22)
β_1 (% Minority)	-2.33	-3.02	-2.29	-3.01	-3.14
	(0.51)	(0.68)	(0.20)	(0.59)	(0.99)
β_2 (% High School)	0.13	0.16	0.10	0.15	0.17
	(0.04)	(0.12)	(0.13)	(0.08)	(0.12)
β_3 (Logged Mean Income)	-0.38	-0.38	-0.34	-0.38	-0.37
	(0.21)	(0.18)	(0.12)	(0.11)	(0.14)
β_4 (% Urban)	-0.68	-1.24	-0.77	-1.24	-1.27
	(0.29)	(1.17)	(0.20)	(0.88)	(0.53)
β_5 (Population Density)	-0.53	-0.61	-1.24	-0.48	-0.58
	(0.29)	(0.29)	(0.11)	(0.28)	(0.28)
γ_1 (Weight Dry)	0.54	0.65	0.59	0.65	0.65
	(0.12)	(0.08)	(0.03)	(0.08)	(0.05)
σ_1^u (Variance Dry)	0.63	0.79	0.73	0.80	0.79
	(0.11)	(0.10)	(0.11)	(0.10)	(0.09)
σ_2^u (Variance Wet)	0.23	0.23	0.24	0.23	0.23
	(0.16)	(0.11)	(0.02)	(0.29)	(0.16)
			Referenda		
g_{1s} (Status Quo)	–	0.31	0.26	0.01	0.01

	(1)	(2)	(3)	(4)	(5)
		(0.18)	(0.06)	(0.25)	(0.21)
g_{1r} (Alternative)	–	-0.74	-0.59	-0.54	-0.69
		(0.12)	(0.08)	(0.08)	(0.09)
σ_2^{ϵ} (Variance Wet)	–	1.15	1.16	1.49	1.33
		(0.46)	(0.11)	(0.13)	(0.09)
Number of Representatives	173	173	226	173	173
Number of Referenda	–	122	122	81	90
Cor(g_1, $\widehat{g_1}$)	0.57	0.54	0.64	0.53	0.54
Cor(V, $\widehat{V}$)	–	0.56	0.60	0.45	0.35
	(1)	(2)	(3)	(4)	(5)

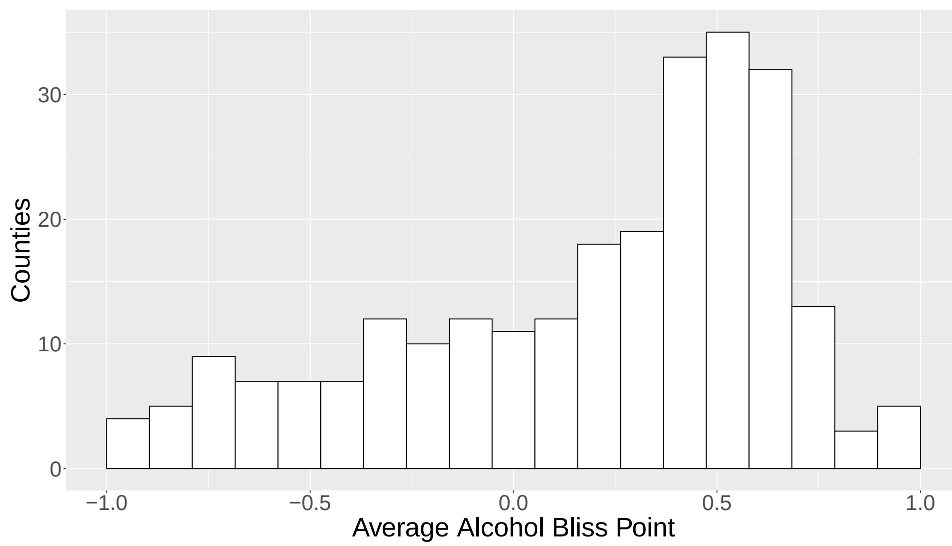
We find that our estimates and standard errors are plausible. We allow for type-specific intercepts in all specifications but constrain the parameters of the observed socio-economic characteristics to be the same across types. The estimates of the α_1 and α_2 reflect the main differences between the two types of voters. Not surprisingly, there are large differences in preferences between dry and wet religious types. Almost all the estimates of β's that capture heterogeneity across districts have the expected sign. Minorities and higher-income voters as well as voters from more densely populated districts prefer wet policies. These findings are similar to the ones that can be obtained from reduced-form regression estimates.[25] The parameter γ_1 captures differences in preference weights and the mobility of the swing voter of the two types. We find that the estimates of this parameter range between 0.54 and 0.65. This suggests that politicians assign a similar weight to both voter types. Dry voters tend to have a slightly larger weight than wet voters.

Figure 3 plots histograms for the predicted bliss points for liquor policy. We find that

[25] Details of the reduced-form estimates are available upon request from the authors.

there is a large amount of heterogeneity in average bliss points. This heterogeneity is driven by differences in socio-economic characteristics within counties, such as differences in income, age, or minority composition, as we have seen in Table 4.

Figure 3: Histogram of Bliss Points



Finally, the estimates of the parameters that are identified by the referenda orthogonality conditions suggest that the position of the dry status quo is positive, reflecting the preferences during the Prohibition Era. The alternative policy, which legalizes the purchase and consumption of beer and wine, is negative and therefore closer to the preferences of the wet voters. The variances of the election shocks are close to one. Again, we view these findings as providing strong evidence in support of our modeling approach.

5.2 Robustness Analysis

We have conducted some robustness analyses. First, we can also estimate the bliss points of 53 legislators that come from 12 multimember districts. In theory, all legislators that represent the same district in a given period should have similar bliss points. Column (3) includes the legislators from multi-member districts in the sample. We find that the overall fit of the model improves as we use these additional data points while the parameter estimates are similar to the ones reported in Column (2).

Second, we have noted that the period in which legislators voted on liquor policies and the period for which we observe referenda do not perfectly overlap. Not surprisingly, legislators' voting typically precedes referenda. We have estimated the model using referenda that took place before the U.S. entry into World War II. One may conjecture that liquor preferences may have changed during and after World War II. Hence, specification (4) only uses referenda that took place before 1942. Overall, the results are similar across all these specifications, which is consistent with the observation that alcohol consumption was very stable between 1935 and 1955. There are no obvious structural changes as can be seen from Figure 1.

Third, Column (5) excludes any referenda that did not pit "beer only" against dry. Overall, we find that our point estimates are similar to the ones of our preferred specification. In particular, there are no large differences in the estimates of the position of the alternative policy. That suggests that any type of "wet" policy that was considered by voters was pushing the policy significantly to the left of the policy space. Again, this finding is consistent with the observation that Prohibition primarily reduced the consumption of beer. Lifting the restrictions on beer consumption thus removed the most obvious constraint imposed by Prohibition policies.

Fourth, the estimates reported in Table 4 are based on an unweighted sample. We have some missing observations in our sample. First, there are some legislators and hence districts for which we cannot estimate the bliss points. More importantly, there are a larger number of counties that never held a single liquor referendum. To address these issues, we have explored different weighting schemes (such as inverse probability weighting) both for the sample of legislators and the sample of referenda. Overall, our main findings are qualitatively and quantitatively robust to these changes.

Finally, we consider how robust our analysis is to assumptions regarding voter turnout. Thus far we have assumed that all citizens turn out to vote in public referenda. It is straightforward to extend the model and allow for differences in voter turnout among the types of voters. Let us assume that the exogenous probability of voter turnout during a referendum is p_i. Define the effective or turnout adjusted population as

$$s_i^e = \frac{p_i \; s_i}{\sum_{j=1}^{J} p_j \; s_j} \tag{25}$$

Hence the effective vote share in favor of the referendum is:

$$V(g_{ks}, g_{kr}) = \sum_i s_i^e \; F_{ki}((g_{ks} - \theta_{ki})^2 - (g_{kr} - \theta_{ki})^2) \tag{26}$$

The estimates reported in Table 4 assume that the voter turnout is the same for the two types ($p_1 = p_2$). This assumption then implies that $s_i^e = s_i$ for all i. Note that we do not have to assume that turnout is complete ($p_1 = 1 = p_2$), which is not the case in these local elections. We just have not allowed for differential attrition in turnout by type ($p_1 \neq p_2$).

We have conducted a more systematic analysis of voter turnout in our sample. The results are summarized in Table 6 in Appendix C. The table contains the estimates of models that regress turnout on the share of voter types across counties. Overall, we find that the coefficient of the share of dry religion voters is insignificant in the regression model. We thus do not find any evidence that suggests differential voter turnout among the two religious types in our model.[26]

[26] Coate and Conlin (2004) and Coate, Conlin, and Moro (2008) provide a more systematic analysis of voter turnout in these local referenda. They also provide some tests of structural turnout models.

5.3 Goodness of Fit

First, we consider the within-sample fit of our model. Figure 4 plots the estimated and predicted bliss points for legislators in the House of Representatives. These estimates are based on the second specification in Table 4. Overall, the within-sample fit of our model is quite good. As shown in Table 4, the correlations between the predicted and estimated bliss points exceed 0.54. As one would expect the fit is not as good for the bliss points of the more extreme politicians. Our model tends to over-predict the bliss points for extreme wet politicians and under-predict the position of extreme dry politicians.

Figure 4: Liquor Policy Bliss Points: Fit

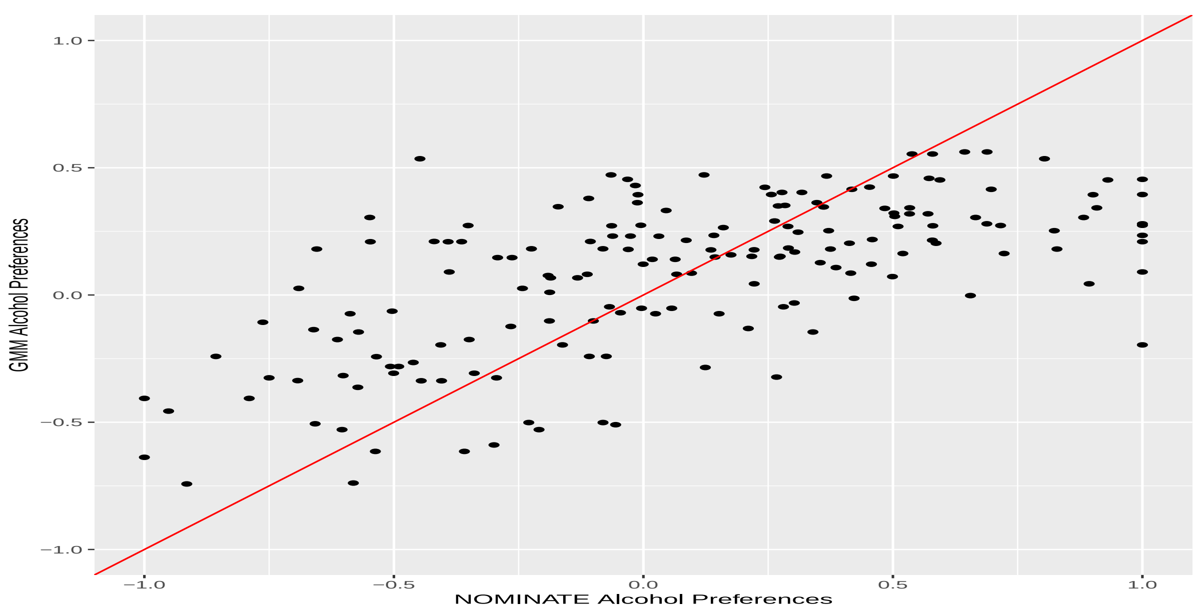
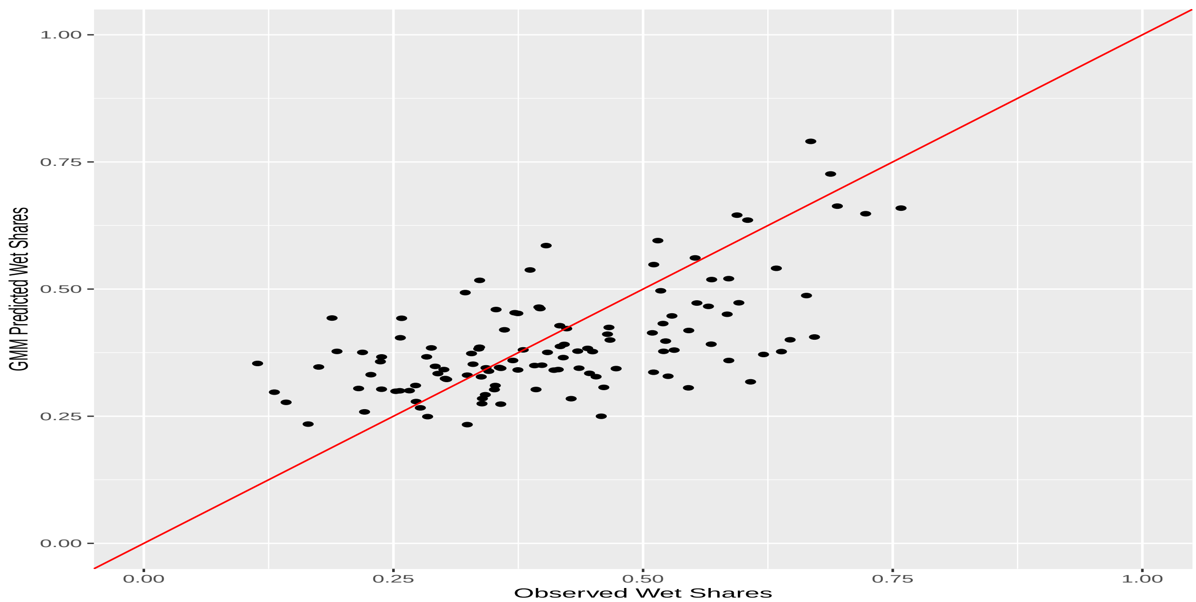


Figure 4 plots the estimated and predicted average vote shares. These estimates are also based on the second specification. We find that the correlation between the predicted and estimated vote shares is approximately 0.56. Overall, the fit is quite good.

To validate our model, we consider the 30 members of the Texas Senate as a hold-out sample. We have also estimated the liquor policy bliss points for these Senators. Note

Figure 5: Wet Share Fit



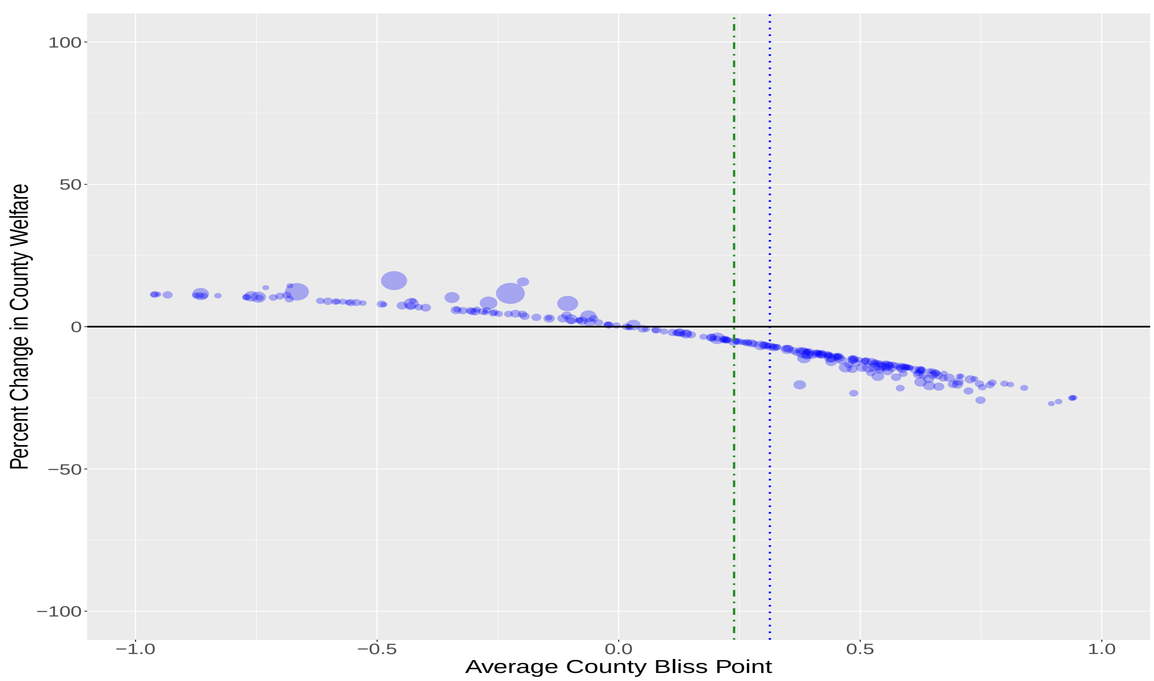
that we did not use these Senators when we estimated the parameters of our voting model. Nevertheless, we can compare the estimated bliss points with the bliss points predicted by our model. We find that the correlation between the estimated and predicted bliss points is 0.54. Our model, therefore, performs as well in predicting the bliss points of the senators as it does for members of the House of Representatives. We thus conclude that we obtain good out-of-sample predictions for the Senate, even though the Senate was more polarized on liquor policy than the House of Representatives. We view these results as providing strong support for the validity of our model.

6 Policy and Welfare Analysis

6.1 Optimal Decentralization

Having estimated the parameters of our model and the position of the Prohibition Era status quo, we can now evaluate the benefits of adopting an optimal centralized or decentralized policy. Here we use the full set of 254 counties in Texas as the relevant set of jurisdictions. Implementing the welfare analysis, we find that the optimal uniform policy is fairly close to the Prohibition Era status quo, $g_u = 0.23 < 0.31 = g_s$. This difference is primarily due to the fact that the counties with large, urban populations tend to prefer a wet policy. Compared to the Prohibition Era status quo, we find that aggregate welfare increases by only 1 percent under the optimal uniform centralized policy.

Figure 6: Status Quo versus Optimal Uniform Policy



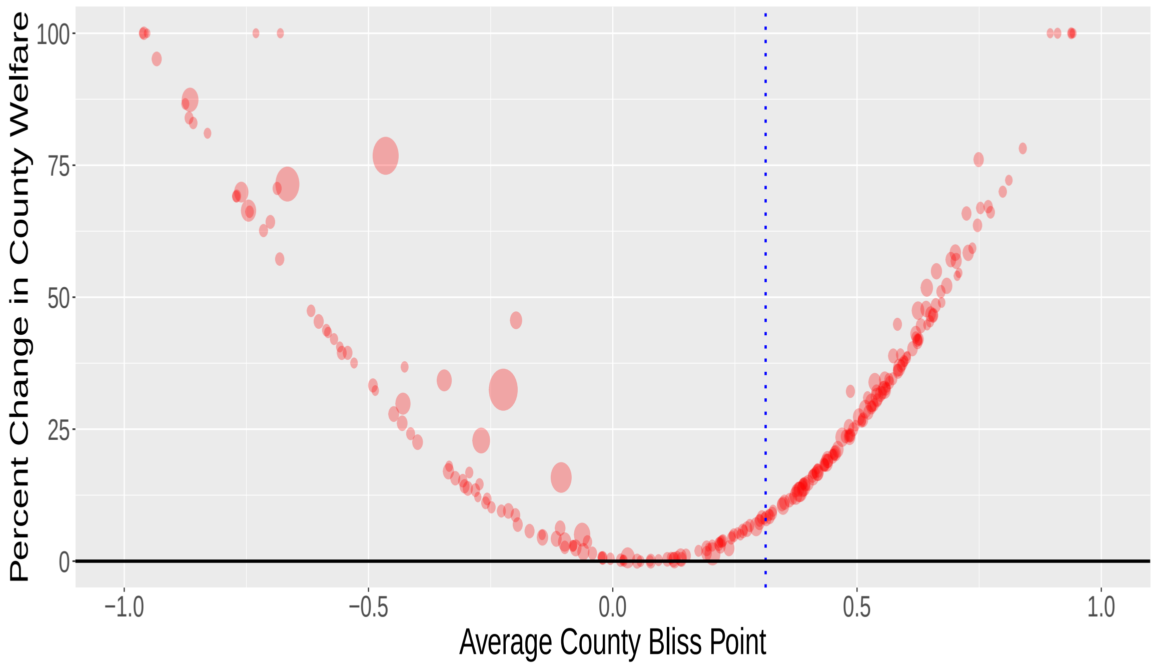
We can also provide a more disaggregated analysis of the welfare gains. In particular, we study which counties gain and which counties lose in the different policy regimes. We plot a county's change in welfare relative to the Prohibition Era status quo against its average alcohol bliss point in Figure 6. Each blue circle represents a county in Texas. The size of the circle is proportional to its voting-age population, so that larger counties are represented with larger circles.

Figure 6 shows the change in welfare comparing the status quo with the optimal uniform centralized policy. As a consequence of this policy shift, counties with average bliss points up to approximately 0.15 experience welfare gains, while counties with bliss points below that threshold experience losses. Also, note that the welfare losses for some small dry counties are quite large. We thus conclude that by the end of the Prohibition Era, the dry status quo was fairly close to the optimal uniform policy.

Our analysis also reveals that there are large gains from decentralization. Compared to the Prohibition status quo, welfare increases by 36.1 percent under the optimal decentralized policy. Compared to the optimal uniform centralized policy, the optimal decentralized policy increases aggregate welfare by 35.5 percent.

Figure 7 plots the change in welfare comparing the status quo with the optimal decentralized policy. Note that all communities gain from policy decentralization. The gains are increasing as we move into the tails of the distribution, which are comprised of counties that have strong preferences for or against liquor sales. We conclude that the compromise that was reached after Prohibition significantly improved welfare. There are large and significant gains from decentralization. These gains are likely to be larger than the costs associated with decentralization. Our findings, therefore, provide strong evidence supporting the decision of the Texas legislature to decentralize liquor policies after the end of the Prohibition Era.

Figure 7: Status Quo versus Optimal Decentralized Policy



6.2 Discussion

To gain some additional insights into the impact of liquor policies on economic welfare, it is useful to map the policy position from the ideological policy space, which is defined on the interval $[-1, 1]$, into the space of average alcohol consumption. In this section, we construct this mapping and illustrate how to construct conventional welfare measures that are associated with the policy changes.

Recall that g denoted the position of the policy in the ideological space. Let $a(g)$ denote the corresponding level of average alcohol consumption as a function of g. It is plausible to assume that the most extreme dry policy position, denoted by $g = 1$, is associated with complete abstinence, i.e. $a(1) = 0$. The estimated average consumption during the period 1920-1921 was 0.33 gallons of pure ethanol per capita. This suggests that the policy had been shifted very close to the extreme dry policy during the peak of Prohibition that saw the most serious attempts to enforce the Volstead Act. For the sake of concreteness, let us assume that $a(0.9) = 0.33$.

By the mid-1920s enforcement efforts had considerably slackened. Alcohol consumption in 1927-1930 averaged around 1.61 gallons of pure ethanol per capita. Similar levels for consumption were maintained from 1931-1933 according to estimates of Miron and Zwiebel (1991). According to Table 4, the average estimate of the dry status quo is approximately 0.15. We, therefore, assume that $a(0.15) = 1.6$. We have also seen in Figure 1 that average alcohol consumption never exceeded 2.5 gallons of pure alcohol per capita between 1900 and 1955. As a consequence, we set $a(-1) = 2.5$. Recall that the average consumption in 1911-14 was approximately 2.39 gallons, which is fairly close to the most extreme wet position. Hence, we assume that $a(-0.9) = 2.39$. After the end of Prohibition, average alcohol consumption rose to 1.85 gallons of pure alcohol from 1935-1955. Our average estimate of the alternative referenda position in Table 4 is approximately -0.64 which suggests that $a(-0.64) = 1.85$.

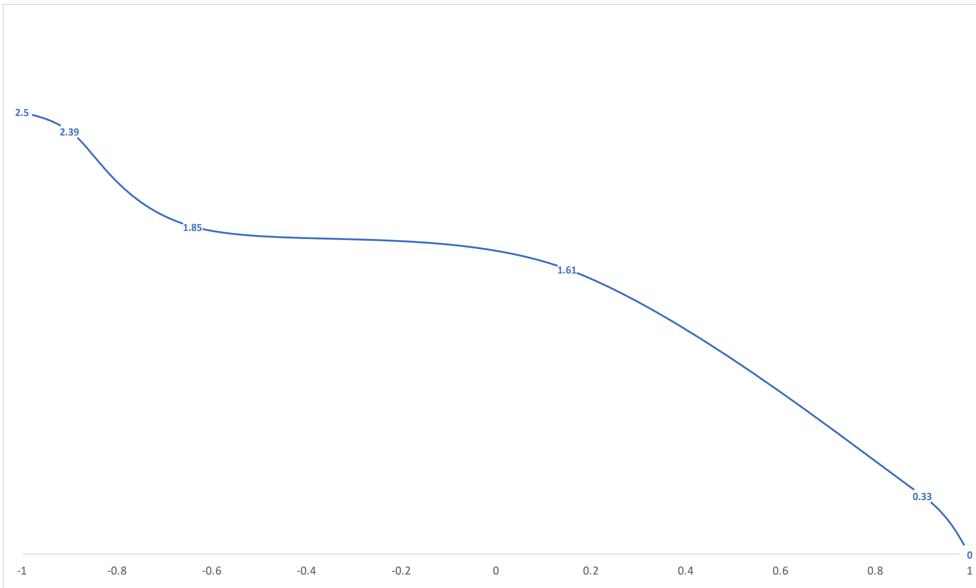
Figure 8 illustrates our mapping from our ideological policy space into average levels of pure alcohol consumption. Note that the mapping is nonlinear, especially for extreme policies.

Next, we illustrate how to construct traditional, dollar-based welfare measures of the benefits of decentralization. Admittedly, the quality of the historical data could be much better. Nevertheless, it may be useful to produce some back-of-the-envelope calculations to get a sense of the magnitude of the gains associated with decentralization. Here we consider the problem of measuring consumer surplus. Given a point $(q_0, P(q_0))$ on the inverse demand curve, consumer surplus is defined as

$$CS(q_0) = \int_0^{q_0} [P(q) - P(q_0)]\, dq \qquad (27)$$

To compute this welfare measure, we need to approximate demand functions for alco-

Figure 8: Mapping the Policy Space into Average Pure Alcohol Consumption Levels



hol. This is admittedly rather challenging, since the quality of price data is even more problematic than the quantity data during the Prohibition Era, as discussed in detail in Warburton (1932). In Appendices D and E, we discuss in detail how to use the data in Warburton (1932) to estimate an iso-elastic demand function for pure alcohol during the Prohibition Era. Our preferred IV estimate of the price elasticity is approximately -1.62, which is similar to the point estimate of -1.8 reported in Cook and Tauchen (1982). Their estimate is based on the post-Prohibition Era and obtained using a much larger data set. Plugging our estimated inverse demand function into equation (27), we can approximate the consumer surplus for all policies. Table 5 summarizes our estimates for four points of the policy space.

According to Warburton (1932), prices per gallon of ethanol ranged between \$24.82 in 1920 and \$10.01 in 1930. This implies that average expenditures were approximately

Table 5: Estimates of Consumer Surplus

Time Period	Alcohol Consumption per capita	Consumer Surplus
1911 - 1914	2.39	\$30.17
1920 - 1921	0.33	\$14.12
1927 - 1930	1.61	\$25.93
1935 - 1955	1.85	\$27.35

\$8.69 in 1920 and \$14.60 in 1930. Relative to these estimates of average expenditures, our estimates of the consumer surplus are large and economically meaningful. We thus conclude that optimal decentralization is likely to be associated with significant gains in consumer surplus.

Consumer surplus measures only one aspect of the economic welfare associated with alternative liquor policy options. For example, it does not capture the aversion to alcohol consumption which characterizes the position of those who prefer to completely abstain from alcohol consumption. A more comprehensive analysis of the monetary benefits of various policies also needs to take into consideration the impact of the different policies on tax revenues. Warburton (1932) estimated that the federal government lost approximately one billion dollars of tax revenues per year during the Prohibition Era. Finally, one needs to measure producer surplus. In the case of Prohibition, a large fraction of the producer surplus was absorbed by organized crime, which was particularly problematic from a welfare perspective. We acknowledge that these aspects were of paramount importance in the discussion to end Prohibition. While our estimates of the ideological policy position capture all these dimensions, including some of the potential costs associated enforcement, it is rather difficult to assign monetary values to all the relevant components that affect the distribution of economic welfare among the citizens.

7 Conclusions

We have studied the most famous natural experiment ever conducted regarding policy decentralization in the U.S., namely the decentralization of liquor policies at the end of the Prohibition Era. We have provided new estimates of the benefits of policy decentralization. Compared to the Prohibition Era status quo, we find that aggregate welfare increases by 1 percent under the optimal uniform centralized policy. Decentralized policies offer opportunities to significantly increase welfare. Compared to the optimal uniform centralized policy, the optimal decentralized policy increases aggregate welfare by 36 percent. Finally, we have documented that there is large heterogeneity in preferences over liquor policies across voting districts and counties in our sample. Our empirical results thus provide a clear interpretation of the failures of the Prohibition Era. We illustrate the usefulness of policy decentralization and provide plausible estimates of the magnitudes of the welfare gains that can arise from decentralization.[27]

The paper has also made some important methodological contributions that go beyond the study of fiscal decentralization. In particular, we have shown how to identify and estimate a probabilistic voting model using (estimated) bliss points of legislators as well as characteristics of voting districts and voter types. We have shown that these models can be estimated for arbitrarily large policy spaces. The proof of identification is constructive and can be used to derive a GMM estimator for the parameters of the model. To estimate this model we do not need to observe public referenda.

We have also shown how to integrate additional data on local referenda in the estimation procedure. If we observe vote shares in local referenda, we can construct additional

[27] There are other fruitful applications which combine legislative voting with local public referenda. For example, Masket and Noel (2012) consider voting on budgetary and educational policy issues in the California legislature as well as outcomes of local public expenditure referenda. Our model is well-suited to study these topics.

orthogonality conditions which help improve the efficiency of the estimator and provide estimates of the position of the status quo and the alternative policy considered in the referenda. These methodological insights should provide ample scope for future research. For example, we could use the same model to study decision-making in cities and assess the impact of heterogeneity within cities on policy outcomes.

Finally, we have shown how to estimate a model with multiple policy dimensions. Primarily due to data limitations, we restricted attention to the case with additively separable preferences, which then allows us to study each policy dimension in isolation. In principle, it is straightforward to extend our estimator to model with non-separable preferences as long as one can estimate all dimensions of the preferences of legislators based on roll call votes. In practice, this might be rather challenging. Future research should investigate these issues more carefully.

We acknowledge that we do not provide a complete cost-benefit analysis of policy decentralization. Instead, we only focus on the voters' benefits. In particular, we do not attempt to estimate the heterogeneity in costs across jurisdictions and voting districts. Heterogeneity in costs could arise, for example, due to different enforcement technologies used at the local level.

Similarly, our analysis should be extended to account for policy spillovers as discussed, for example, in Acemoglu, Garcia-Jimeno, and Robinson (2015). In our application, spillover effects may arise due to liquor purchases by customers that cross jurisdictional borders or due to spillovers in law enforcement. Moreover, drunk driving and highway accidents may cross county boundaries. Finally, individuals who are against liquor consumption plausibly get disutility from the very fact that others are drinking even if they live in another jurisdiction. This is just one example of peer effects that may be important in this context. It is possible that preferences of liquor policies in one county depend on preferences of neighboring counties and are defined over policy outcomes in neighboring jurisdictions. There is very little empirical evidence on the magnitude of these spillover effects during the post-Prohibition Era. Future research might fruitfully investigate the feasibility of incorporating of spillover and peer effects into the analysis.

References

Acemoglu, D., Garcia-Jimeno, C. and J. Robinson (2015). State Capacity and Economic Development: A Network Approach. American Economic Review, 105(8), 2364-2409.

Aruoba, B., Drazen, A., and Vlaicu, R. (2019). A Structural Model of Electoral Accountability. International Economic Review, 60 (2), 517-45.

Battaglini, M., Patacchini, E. and E. Rainone (2020). Endogenous Social Connections in Legislatures, Working Paper.

Besley, T. and S. Coate (1997). An Economic Model of Representative Democracy. The Quarterly Journal of Economics. 112 (1), 85-114.

Besley, T. and S. Coate (2003). Centralized versus Decentralized Provision Local Public Goods: A Political Economy Approach. Journal of Public Economics, 87, 2611-37.

Calabrese, S. (2000). Multimember District Congressional Elections. Legislative Studies Quarterly. 25 (4), 611-643.

Calabrese, S., Epple, D., and R. Romano (2012). Inefficiencies from Metropolitan Political and Fiscal Decentralization: Failures of Tiebout Competition. Review of Economic Studies 79(3), 1081-1111.

Calabrese, S., Epple, D., and R. Romer, T. and H. Sieg (2006), Local Public Good Provision: Peer Effects, Voting, and Mobility. Journal of Public Economics, 90 (6-7), 959-981.

Clinton, J., Jackman, S., and D. Rivers (2004). The Statistical Analysis of Roll-call Data. American Political Science Review 98, 355-70.

Coate, S. and M. Conlin (2004). A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence. American Economic Review. 94 (5), 1476-1504.

Coate, S., Conlin, M., and Moro, A. (2008). The Performance of Pivotal-Voter Models in Small-Scale Elections: Evidence from Texas Liquor Referenda. Journal of Public Economics, 92, 582-96.

Cook, P. and G. Tauchen (1982). The Effect of Liquor Taxes on Heavy Drinking, Bell Journal of Economics, 13, 379-390.

Coughlin, P. (1992). Probabilistic Voting Theory. Cambridge University Press.

Degan, A. and A. Merlo (2011). A Structural Model of Turnout and Voting in Multiple Elections. Journal of the European Economic Association, 9 (2), 209- 245.

Diermeier, D., Eraslan, H. and A. Merlo (2003). A Structural Model of Government Formation. Econometrica, 71, 27-70.

Eckham, D. (2016). Multimember District Representation in the US and in the States. Web Report.

Epple, D., Romer, T., and H. Sieg (2001). Inter-jurisdictional Sorting and Majority Rule: An Empirical Analysis. Econometrica, 2001, 69 (6), 1437-1466.

Fowler, Anthony and Hall, Andrew (2012), Conservative Vote Probabilities: An Easier Method for the Analysis of Roll Call Data, Working Paper.

Garcia-Jimeno, C. (2016). The Political Economy of Moral Conflict: An Empirical Study of Learning and Law Enforcement Under Prohibition. Econometrica, 84(2), 511-570.

Gacia-Jimeno, C, Iglesias, A. and P. Yildirim (2021). "Information Networks and Collective Action: Evidence from the Women's Temperance Crusade." The American Economic Review, Vol. 112, No. 1, 41-80.

Heckman, J. and J. Snyder (1997). The RAND Journal of Economics. Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators, Special Issue in Honor of Richard E. Quandt, 28 (0), S142-S189.

Iaryczower, M. and M. Shum (2012). The Value of Information in the Court: Get it Right, Keep it Tight, American Economic Review, 102, 202-37.

Kawai, K. and Y. Watanabe (2013). Inferring Strategic Voting. The American Economic Review, 103 (2), 624-662.

Knight, B. and N. Schiff (2010). Momentum and Social Learning in Presidential Primaries. Journal of Political Economy, 118 (6), 1110-1150.

Lindbeck, A. and J. Weibull (1987). Balanced-budget redistribution as the outcome of political competition. Public Choice, 52(3), 273-297.

Lockwood, B. (2002). Distributive Politics and the Benefits from Decentralization. Review of Economic Studies 69 (2), 313-338.

Lott, J., and L. Kenny (1999). Did Women's Suffrage Change the Size and Scope of Government? Journal of Political Economy, 107 (6), 1163-1198.

Masket, S. and H. Noel (2012). Serving Two Masters: Using Referenda to Assess Partisan versus Dyadic Legislative Representation. Political Research Quarterly, 65 (1), 104-23.

McCarty, N., Poole, K., and H. Rosenthal (2006). Polarized America: The Dance of Ideology and Unequal Riches. Cambridge, MA: MIT Press.

Merlo, A. (1997). Bargaining over Governments in a Stochastic Environment. Journal of Political Economy, 105, 101-131.

Merlo, A. and X. Tang (2012). Identification and Estimation of Stochastic Sequential Bargaining Games. Econometrica, 80, 373-389.

Miller, G. (2008). Women's Suffrage, Political Responsiveness, and Child Survival in American History. The Quarterly Journal of Economics, 123 (3), 1287-1327.

Miron, J., and J. Zwiebel, Alcohol Consumption During Prohibition, American Economic Review, May 1991, 81, 242-47.

Newey, W. and McFadden, D. (1994) Large Sample Estimation and Hypothesis Testing. In: Engle, R. and McFadden, D., Eds., Handbook of Econometrics, Elsevier, 2111-2245.

Oates, W. (1972). Fiscal Federalism. Horcourt Brace. New York.

Osborne, M. and A. Slivinski (1996). A Model of Political Competition with Citizen-Candidates. The Quarterly Journal of Economics, 111 (1), pp. 65-96.

Poole, K. and H. Rosenthal (1985). A Spatial Model for Legislative Roll Call Analysis. American Journal of Political Science. 29 (2), 357-384.

Poole, K. and H. Rosenthal (1991). Patterns of Congressional Voting. American Journal of Political Science, 35, 228- 278.

Romer, T. and H. Rosenthal (1978). Political Resource Allocation, Controlled Agendas, and the Status Quo. Public Choice, 33 (4), 27-43.

Sieg, H. and C. Yoon (2017). Estimating Dynamic Games of Electoral Competition to Evaluate Term Limits in U.S. Gubernatorial Elections. American Economic Review, 107 (7), 1824-57.

Sieg, H. and C. Yoon (2022). Electoral Accountability and Control in U.S. Cities. Journal of Political Economy, forthcoming.

Strumpf, K. and F. Oberholzer-Gee (2002). Endogenous Policy Decentralization: Testing the Central Tenet of Economic Federalism. Journal of Political Economy. 110 (1), 1-36.

Warburton, C. (1932). The Economic Results of Prohibition. Columbia University Press. New York.

A Identification: Proof of Proposition 1

For expositional simplicity consider the case in which $I = 2$. In the spirit of an identification at infinity result, we consider two extreme cases of the model which allows us to establish identification of (α, β). First, note that if $s_{1n} = 1$, our model reduces to the following simple regression model:

$$g_{kn} \;=\; \alpha_{k1} + \beta_{k1}\, x_{kn} \;+\; u_{k1n} \tag{28}$$

Hence $(\alpha_{k1}, \beta_{k1})$ are identified. This also implies that the conditional distribution of u_{k1n} is non-parametrically identified. So we do not need a distributional assumption for the error terms to establish identification. Similarly, if $s_{1n} = 0$ and hence $s_{2n} = 1$, we have:

$$g_{kn} \;=\; \alpha_{k2} + \beta_{k2}\, x_{kn} \;+\; u_{k2n} \tag{29}$$

Hence $(\alpha_{k2}, \beta_{k2})$ are identified. This argument easily extends to the more general case in which to $I > 2$.

Next, define the expected bliss points for each type k as:

$$\begin{aligned}\bar{\theta}_{ikn} &= E[\theta_{kin}|x_{kn}] \qquad (30)\\ &= \alpha_{ki} + \beta_{ki}\, x_{kn}\end{aligned}$$

which we can treat as known, since α and β are identified. Now consider the case in which $0 < s_{in} < 1$. We can write g_{kn} as a weighted average of the expected bliss points and the error term:

$$g_{kn} \;=\; \sum_i \frac{s_{in}\gamma_{ki}}{\sum_j s_{jn}\gamma_{kj}}\, \bar{\theta}_{ikn} \;+\; v_{kn} \tag{31}$$

The easiest way to see that the γ_{ki}'s are only identified up to a normalization is to consider an imposture structure $c\,\gamma_{ki}$, where c is an arbitrary positive constant. For any value of

s_{1n} we have:

$$\frac{s_{in}c\ \gamma_{ki}}{\sum_i s_{in}c\ \gamma_{ki}} \quad = \quad \frac{c\ s_{in}\gamma_{ki}}{c\ \sum_i s_{in}\gamma_{ki}} = \frac{s_{in}\gamma_{ki}}{\sum_i s_{in}\gamma_{ki}} \tag{32}$$

To put it differently, the weights above must sum up to one. A natural normalization is:

$$\sum_{i=1}^{I} \gamma_{ki} = 1 \qquad k = 1, 2 \tag{33}$$

Additional restrictions can be obtained if impose some assumptions on the distribution of the error terms, For example, if $V(u_{kin}) = \sigma_{ik}^2$, then the conditional variance of the error terms is given by:

$$Var(v_{kn}|\ s_n) \quad = \quad \sum_i \left(\frac{s_{in}\gamma_{ki}}{\sum_j s_{jn}\gamma_{kj}} \right)^2 \sigma_{ki}^2 \tag{34}$$

Hence the σ_{ki} are identified from the equation above. Q.E.D.

B GMM Estimation

Notet hat

$$\theta_{kni} = \alpha_{ki} + \beta_k x_{kn} + \epsilon_{kni} \tag{35}$$

We assume that

$$\theta_{kni} \sim N(\mu_{kni}, \sigma_{ki}^2, -1, 1) \tag{36}$$

where $\mu_{kni} = \alpha_{ki} + \beta_k x_{kn}$. Hence, we have that

$$\epsilon_{kni} \ \sim \ N(0, \sigma_{ki}^2, -1 - \mu_{kni}, 1 + \mu_{kni}) \tag{37}$$

Define $l_{kni} = (-1 - \mu_{kni})/\sigma_{ki}$ and $u_{kni} = (1 - \mu_{kni})/\sigma_{ki}$ As a consequence we have

$$E[\theta_{kni}|x_{kn}] \quad = \quad \alpha_{ki} + \beta_k x_{kn} - \sigma_{ki} \left(\frac{\phi(u_{kni}) - \phi(l_{kni})}{\Phi(u_{kni}) - \Phi(l_{kni})} \right) \tag{38}$$

and

$$\sigma^2_{kni} \;=\; \sigma^2_{ki}\left[1 - \frac{u_{kni}\,\phi(u_{kni}) \;-\; l_{kni}\,\phi(l_{kni})}{\Phi(u_{kni}) - \Phi(l_{kni})} - \left(\frac{\phi(u_{kni}) - \phi(l_{kni})}{\Phi(u_{kni}) - \Phi(l_{kni})}\right)^2\right] \tag{39}$$

Because of the separability of the utility function, we can estimate the model separately for each policy dimension. To simplify the notation, we consider again the case in which $I = 2$.[28] As we discussed in Section 4 of the paper, we can derive orthogonality conditions from a variety of different implications of the model. First, consider the predictions of our model for the bliss points of legislators. For each policy dimension k define a parameter vector:

$$\phi_{1k} \;=\; (\gamma_{k1}, \alpha_{k1}, \beta_{k1}, \alpha_{k2}, \beta_{k2}) \tag{40}$$

For policy dimension k , we obtain the following orthogonality condition:

$$h_{1k}(g_{kn}, s_{1n}, s_{2n}, x_{kn}|\phi_{1k}) = g_{kn} - \sum_i \frac{\gamma_{k1}\; s_{in} E[\theta_{kni}|x_{kn}]}{\sum_j s_{jn}\gamma_{kj}} \tag{41}$$

Next consider the variance of the error terms. For each k, define a second parameter vector:

$$\phi_{2k} \;=\; (\sigma_{k1}, \sigma_{k2}) \tag{42}$$

Define the residuals for policy dimension k as:

$$v_{kn} \;=\; h_{1k}(g_{kn}, s_{1n}, s_{2n}, x_{kn}|\; \phi_{1k}) \tag{43}$$

Based on the variance of the error terms, we can define the following orthogonality condition:

$$h_{2k}(v^2_{kn}, s_{1n}, s_{2n}|\; \phi_{1k}, \phi_{2k}) \;=\; v^2_{kn} - Var[v_{kn}|s_{1n}, s_{2n}, \phi_{1k}, \phi_{2k}] \tag{44}$$

[28]The estimator easily generalizes to the more general case, when $I \geq 2$.

where

$$Var[v_{kn}|s_{1n}, s_{2n}, \phi_{1k}, \phi_{2k}] \;=\; \sum_i \left(\frac{s_{in}\gamma_{ki}}{s_{1n}\gamma_{ki} + s_{2n}(1-\gamma_{k1})} \right)^2 \sigma^2_{kni} \tag{45}$$

Note that we are exploiting the condition here that the functional form of the heteroskedasticity is known.

Finally consider the vote shares in referenda that pit the status quo against the same alternative.. Define a third parameter vector as:

$$\phi_{3k} \;=\; (g_{ks}, g_{kr}, \sigma^{\epsilon}_{k1}, \sigma^{\epsilon}_{k2}) \tag{46}$$

We can exploit the predictions of the vote share in the referenda to obtain the following orthogonality condition:

$$h_{3k}(V_{kn}, s_{1n}, s_{2n}, x_{1n}|\phi_{1k}, \phi_{2k}, \phi_{3k}) \;=\; V_{kn} \;-\; E[V_n|s_{1n}, s_{2n}, x_{kn}] \tag{47}$$

where:

$$E[V_{kn}|s_{1n}, s_{2n}, x_{kn}] = \sum_i s_{in} \int \frac{\exp[(g^2_{ks} + 2\,(g_{kr} - g_{ks})\,(\alpha_{ki} + \beta_{ki}x_{kn}) \;+\; 2\,(g_{kr} - g_{ks})\,u_{ki} \;-\; g^2_{kr})/\sigma^{\epsilon}_{ki}]}{1 + \exp[(g^2_{ks} + 2\,(g_{kr} - g_{ks})\,(\alpha_{ki} + \beta_{ki}x_{kn}) \;+\; 2\,(g_{kr} - g_{ks})\,u_{ki} \;-\; g^2_{kr})/\sigma^{\epsilon}_{k}]}\, r_{ki}(u_{ki})\, du_{ki}$$

For computational convenience, we assume that the $r_{ki}(\cdot)$ are normal densities with mean zero and variance σ^2_{ki}. We can, therefore, numerically integrate the inner summand using quadrature methods. Let ϕ_k denote the full parameter vector:

$$\phi_k \;=\; (\phi_{1k}, \phi_{2k}, \phi_{3k}) \tag{48}$$

We can estimate all parameters jointly using a GMM framework. To see this, define

$$f_{1k}(g_{kn}, s_{1n}, s_{2n}, x_{kn}, z_{1kn}|\phi_k) \;=\; z_{1kn}\; h_{1k}(g_{kn}, s_{1n}, s_{2n}, x_{kn}|\phi_k) \tag{49}$$

where z_{1kn} is a J-dimensional vector of instruments. Note that $J \geq 5$ is a necessary condition for identification. Additional orthogonality conditions can be formed based on the variance restrictions. Similarly, define:

$$f_{2k}(g_{kn}, s_{1n}, s_{2n}, x_{kn}, z_{2kn}|\phi) \;=\; z_{2kn}\; h_{2k}(s_{1n}, s_{2n}|\phi_k) \tag{50}$$

and:

$$f_{3k}(V_{kn}, s_{1n}, s_{2n}, x_{kn}, z_{3kn}|\phi) \quad = \quad z_{3kn}\; h_3(V_{kn}, s_{1n}, s_{2n}, x_{kn}|\phi_k) \tag{51}$$

Define $z'_{kn} = (z'_{1kn}, ...z'_{3kn})$, and $s'_n = (s_{1n}, s_{2n})$. Stacking the orthogonality conditions, we have:

$$f(g_{kn}, V_{kn}, s_n, x_{kn}, z_{kn}|\phi_k) \quad = \quad \begin{pmatrix} f_{1k}(g_{kn}, s_{1n}, s_{2n}, x_{kn}, z_{1kn}|\phi_k) \\ f_{2k}(g_{kn}, s_{1n}, s_{2n}, x_{kn}, z_{2kn}|\phi_k) \\ f_{3k}(V_{kn}, s_{1n}, s_{2n}, x_{kn}, z_{3kn}|\phi_k) \end{pmatrix} \tag{52}$$

and note that

$$E\Big(f(g_{kn}, V_{kn}, s_n, x_{kn}, z_{kn}|\phi^0_k)\Big) = 0 \tag{53}$$

where ϕ^0_k is the true parameter value. The optimally weighted GMM estimator is then defined as:

$$\tilde{\phi}_{kN} = \text{argmin} \ \left[\frac{1}{N}\sum_{n=1}^{N} f(g_{kn}, V_{kn}, s_n, , x_{kn}, z_{kn}|\phi_k)\right]' \ W_{kN} \ \left[\frac{1}{N}\sum_{n=1}^{N} f(g_{kn}, V_{kn}, s_n, x_{kn}, z_{kn}|\phi_k)\right]$$

where W_{kN} is a consistent estimator of the optimal weighting matrix. Note that we can estimate the optimal weighting matrix using the following estimator:

$$W_{kN} = \left[\frac{1}{N}\sum_{n=1}^{N} f(g_{kn}, V_{kn}s_n, x_{kn}, z_{kn}|\hat{\phi}_{kN}) \ f(g_{kn}, V_{kn}s_n, x_{kn}, z_{kn}|\hat{\phi}_{kN})'\right]^{-1} \tag{54}$$

where $\hat{\phi}_{kN}$ is a consistent estimator. A natural starting point are the following instruments:

$$\begin{aligned} z_{1kn} &= (1, s_{1n}, x_{kn}, s_{1n}\; x_{kn}, s_{2n}\; x_{kn}, \frac{1}{s_{1n}}, \frac{1}{s_{2n}})' \\ z_{2kn} &= (1, s_{1n})' \\ z_{3kn} &= (1, x_{1n}, s_{1n}, \frac{1}{s_{1n}}, \frac{1}{s_{2n}})' \end{aligned} \tag{55}$$

C Voter turnout Analysis

Figure 9 shows the distribution of referendum turnout, measured as the total number of votes cast in a county referendum divided by the county's voting age population, in our sample of 302 referenda.

Figure 9: Turnout in 302 Local Referenda

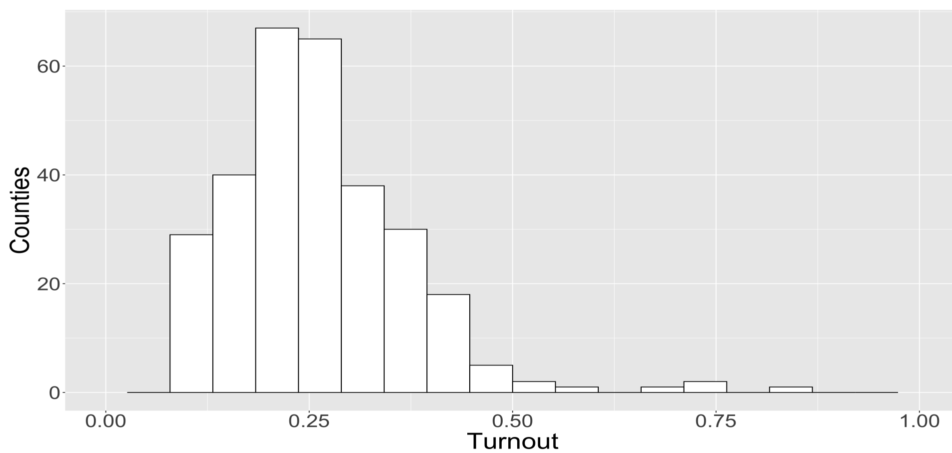


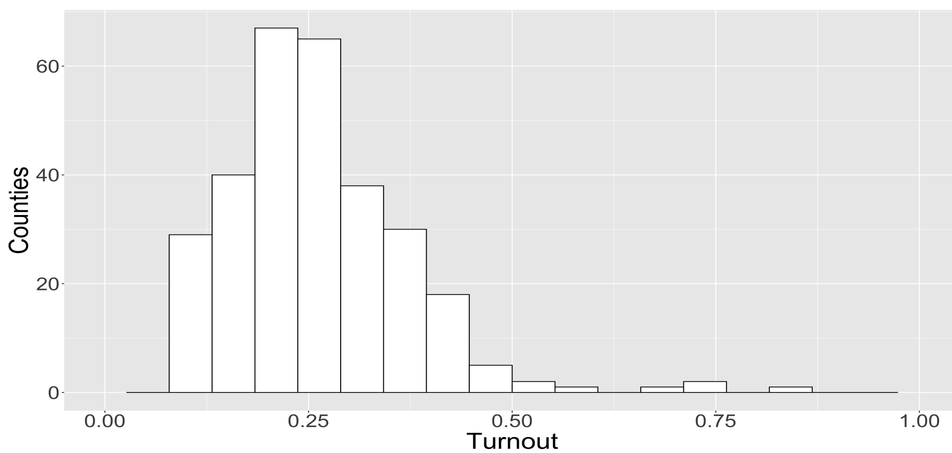
Figure 10 shows the same histogram on the sample of 122 averaged elections that we use in our preferred estimation specification. These statistics are comparable to the ones reported in Coate and Conlin (2004) who analyze turnout during a longer time period than we do.

Table 6 shows the results of regressing turnout on the share of voter types across counties. Columns 1 shows the regression on the full set of 302 elections, and columns 2 repeats the regression on the sample of 122 averaged elections that we use in our preferred estimation specification. The share of dry religion voters in all specifications is insignificant and close to zero. We thus do not find any evidence that suggests differential voter turnout among the two types in our model.

Table 6: Turnout Regressions

	Individual Referenda	Averaged Referenda
% Anti-alcohol Religion	-0.011	-0.047
	(0.042)	(0.050)
% Minority	-0.208***	-0.218***
	(0.054)	(0.067)
% High School	0.145	0.220
	(0.107)	(0.142)
Logged Mean Income	0.064*	0.035
	(0.038)	(0.047)
% Urban	-0.143***	-0.156***
	(0.041)	(0.050)
Population Density (× 100)	-0.016	-0.009
	(0.024)	(0.024)
Observations	302	122
Adjusted R^2	0.165	0.252
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 10: Turnout in 122 Averaged Referenda



D Alcohol Consumption in U.S.

Table 7 reports quantities of alcohol consumed in gallons of pure ethanol per capita. We use total population that is fifteen and older population to obtain the per capita quantities. For the years 1934-1950, quantities come from Slater and Alpert (2021). For the years 1921-1930, quantities are reported in Table 30 of Warburton (1932), Table 30. [29] For the years 1900-1920, quantities come from Warburton (1932), Table 1. Here, Warburton reports per capita quantities with respect to the entire U.S. population of gallons of beer, wine, spirits, and pure ethanol consumed. We convert the gallons of beer, wine, and spirits into gallons of pure ethanol using Warburton's assumed alcohol by volume measures of 4.25 percent, 14 percent, and 50 percent, respectively. We then re-weight the quantities so that they are per capita quantities relative to only the fifteen and older population. Data on alcohol consumption between 1931 and 1933 are based on Miron and Zwiebel (1991).

[29] According to the 1930 Complete Count Census Data, the total U.S. population in 1930 was 122,777,512, of which 86,694,367 individuals were at least fifteen years of age.

Table 7: Per Capita Alcohol Consumption in Gallons of Pure Ethanol, 1900-1955

Year	Beer	Wine	Spirits	Total	Year	Beer	Wine	Spirits	Total
1900	0.97	0.08	0.91	1.95	1928	0.33	0.15	1.21	1.70
1901	0.96	0.07	0.93	1.95	1929	0.40	0.13	1.30	1.83
1902	1.03	0.12	0.95	2.11	1930	0.39	0.13	0.95	1.46
1903	1.06	0.09	1.01	2.17	1931	–	–	–	1.71
1904	1.08	0.10	1.02	2.20	1932	–	–	–	1.44
1905	1.08	0.08	1.00	2.17	1933	–	–	–	1.58
1906	1.17	0.11	1.04	2.32	1934	0.61	0.07	0.29	0.97
1907	1.24	0.13	1.12	2.48	1935	0.68	0.09	0.43	1.20
1908	1.22	0.11	0.98	2.32	1936	0.79	0.12	0.59	1.50
1909	1.15	0.13	0.93	2.21	1937	0.82	0.13	0.64	1.59
1910	1.19	0.13	1.01	2.32	1938	0.75	0.13	0.59	1.47
1911	1.25	0.13	1.03	2.41	1939	0.75	0.14	0.62	1.51
1912	1.20	0.11	1.03	2.35	1940	0.73	0.16	0.67	1.56
1913	1.25	0.11	1.07	2.42	1941	0.81	0.18	0.71	1.70
1914	1.25	0.11	1.02	2.37	1942	0.90	0.22	0.85	1.97
1915	1.11	0.07	0.89	2.07	1943	1.00	0.17	0.66	1.83
1916	1.07	0.09	0.97	2.14	1944	1.13	0.18	0.76	2.07
1917	1.09	0.08	1.15	2.32	1945	1.17	0.20	0.88	2.25
1918	0.90	0.10	0.60	1.60	1946	1.07	0.24	0.99	2.30
1919	0.48	0.10	0.55	1.13	1947	1.11	0.16	0.76	2.03
1920	0.15	0.02	0.18	0.35	1948	1.07	0.20	0.70	1.97
1921	0.06	0.06	0.20	0.31	1949	1.06	0.22	0.70	1.98
1922	0.09	0.07	0.98	1.14	1950	1.04	0.23	0.77	2.04
1923	0.12	0.13	1.27	1.51	1951	1.03	0.20	0.78	2.01
1924	0.15	0.12	1.14	1.41	1952	1.04	0.21	0.73	1.98
1925	0.18	0.12	1.20	1.50	1953	1.04	0.20	0.77	2.01
1926	0.23	0.14	1.29	1.66	1954	1.01	0.21	0.74	1.96
1927	0.28	0.15	1.05	1.48	1955	1.01	0.22	0.77	2.00

E The Demand for Alcohol during Prohibition

To estimate a demand function for alcohol during the Prohibition Era we need to construct price data. Table 78 in Warburton (1932) contains prices per gallon of beer, wine, and spirits from 1920-1930. We convert these prices to prices per gallon of pure ethanol assuming alcohol by volumes of 4.25 percent, 14 percent, and 50 percent for beer, wine, and spirits, respectively. To construct a price of pure ethanol, we compute a weighted sum of the beer, wine, and spirits prices, where the weights on each type of alcohol's price corresponds to that type's share of total alcohol consumption as reported in Table 7.

Table 8: Estimated Demand Schedule for Pure Ethanol, 1920-1930

Year	Per Capita Gallons of Pure Ethanol	Price per Gallon in Dollars
1920	0.35	24.82
1921	0.31	17.72
1922	1.14	13.78
1923	1.51	11.03
1924	1.41	11.09
1925	1.50	11.10
1926	1.66	10.46
1927	1.48	9.92
1928	1.70	11.37
1929	1.83	10.56
1930	1.46	10.01

Table 8 reports the demand schedule for alcohol in the U.S. between 1920 and 1930.

Quantities consumed are reported in gallons of pure ethanol consumed per capita with respect to the fifteen and older population.

We consider an iso-elastic demand curve of the form

$$q \;=\; A\, p^{\epsilon} e^{u} \tag{56}$$

Taking logarithms of equation (56) gives our estimating equation:

$$ln(q) = ln(A) + \epsilon\; ln(p) + u \tag{57}$$

We first estimate this equation using OLS which ignores the potential endogeneity of the price. We also estimate an IV regression using the logged price index of sugar as an instrument.[30] Table 9 summarizes the estimates. Overall, we find that the OLS and IV results are fairly consistent with each other.

[30] The sugar price index comes from Warburton (1932), Table 20, column X_3.

Table 9: Estimating the Price Elasticity of Ethanol Demand, 1920 - 1930

	Dependent variable:	
	log(Ethanol Consumption)	
	OLS	IV
ϵ	-2.018^{***}	-1.622^{***}
	(0.281)	(0.385)
α	5.210^{***}	4.214^{***}
	(0.710)	(0.971)
Observations	11	11
Adjusted R^2	0.835	0.799
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	