**Title**
Social Institutions and Development



**Author**
Egel, Daniel Peter

**Publication Date**
2010

# Social Institutions and Development

by

Daniel Peter Egel

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Barry Eichengreen, Co-Chair
Professor Edward Miguel, Co-Chair
Professor Alain de Janvry
Professor Elisabeth Sadoulet

Spring 2010

Social Institutions and Development

Abstract

Social Institutions and Development
by
Daniel Peter Egel
Doctor of Philosophy in Economics
University of California, Berkeley
Professor Barry Eichengreen, Co-Chair
Professor Edward Miguel, Co-Chair

This dissertation studies the role of social institutions in economic development. While other research has examined the role of ethnicity, religion and other types of large-scale social organizations in development, I study the impact of two different types of local social organizations in two very different contexts. The first social institution that I study are the tribes of modern Yemen and the second is the Freemasons of the United States in the nineteenth century. I demonstrate that both have had an important impact on development, with the first affecting a political patronage network that functions through the education system and the second having a direct impact on the development of the American educational system.

The first chapter examines the relationship between political patronage and tribes, a key social institution in the developing world. Patronage is a tool used throughout the world to reward political allies. Here I create a dataset of Yemeni tribes to explore their role in an educational patronage network that accounts for upwards of 6% of the entire Yemeni government budget. My analysis has two key results. First, conditional on a rich set of controls, I find that the number of tribes has a significant impact on the quantity of patronage. This impact is negative between regions, though positive within regions, as regions with more tribes have less patronage while sub-regions with more tribes have more patronage. The contrast between these effects illustrates the differing influence of tribes in local and national politics. Second, I find no evidence that a recent decentralization reform affected this patronage network. This analysis provides insight into how pre-Islamic institutions have an important role in the development outcomes of the Muslim Middle East and why decentralization reforms proposed for countries similar to Yemen, such as Afghanistan and Somalia, may be ineffective in weakening the power of local elites.

The second chapter examines the role that American Freemasonry played in the historical expansion of the American educational system. I demonstrate that 19th century Freemasonry had a significant positive impact on educational enrollment during and after the rapid rise of the 'common school' in the late 19th

century. And in what is a striking example of the 'path dependence' of social institutions, I show that this effect persisted through the expansion of American high schools in the 1910s-1940s even after the waning of the influence of this organization. Interestingly, Freemasonry's impact was particularly significant in areas that were the most heterogeneous - both ethnically and religiously. This, combined with the the further observation that areas with more Freemasons had higher levels of local taxation, suggests that Freemasonry helped communities overcome the common good problem. As Freemasons did not tend to migrate to areas with existing public education systems, this effect is not driven by reverse causality. And I use a panel data set of enrollment to provide evidence that unobserved heterogeneity and endogeneity are not driving the observed relationship.

The third chapter, which is a co-authored project with Bryan Graham and Cristine Campos de Xavier Pinto, develops a new empirical tool of significant utility for empirical economists studying issues such as those faced in the other chapters. It presents a new estimator, based on minimum empirical discrepancy (MD) methods, for a class of data combination problems. In these problems the researcher does not have access to a random sample containing measurements of all required variables, $Z = (W', X', Y')'$. Instead two separate samples are available. The first is drawn from the *study* population of interest and contains $N_s$ measurements of $(Y, W)$ . The second is drawn from an *auxiliary* population and contains $N_a$ measurements of $(X, W)$ . The first step of our procedure involves using MD methods to re-weight the auxiliary sample in order to match study sample moments of $W$, the variable common to the two datasets. Sample moments from the study and re-weighted auxiliary samples are then combined to estimate the parameter of interest. We show that our estimator's asymptotic variance coincides with the relevant variance bound under two auxiliary parametric restrictions, but only requires one of these two restrictions to hold for consistency ('double robustness'). Our procedure can be used to estimate the average treatment effect on the treated (ATT), the two sample instrumental variables (TSIV) model, counterfactual earnings distributions, and to construct poverty maps. We compare our estimator with leading alternatives in an illustrative study of the effect of National Supported Work (NSW) demonstration participation on earnings and in a series of Monte Carlo experiments.

# Contents

# List of Figures

# List of Tables

## Acknowledgments

I have many people to thank. My father and both of my brothers (Andrea and Paul) have been a tremendous support throughout my academic career and during my six years at Berkeley. And my colleagues, friends and many roommates have made the years enjoyable - and helped me pass the time with board games (i.e. Arkham Horror and Power Grid), breakfast at the faculty club, intramural soccer (both Team Slutsky and the Invisible Hands) and a dozen other distractions. This includes, in alphabetical order for lack of a better option, Raymundo Campos, Jeremiah Dittmar, Rick Fernholz, Brad Howells, Salar Jahedi, Xa Dwight Johnson (the fourth, of course) Andrej Milivojevic, Carlos Noton, Juan Felipe Quintero, Jonathan Rose, Kevin Scallan, Mona Selim el-Sherif, Peter Stirling, Philippe Wingender, and Li Zeng. This thanks also extends to Tanya Goldsmith, my landlady extraordinaire. And, last but certainly not least, Chloe Thurston has been particularly supportive during the last two years of my PhD and read innumerable drafts of my job market paper and helped me survive the job market.

In terms of advising, I have been very lucky to have had the best support that a graduate student could hope for. And while Barry Eichengreen, Bryan Graham and Ted Miguel have been extraordinary in this regard, I have also benefited from the support and encouragement of Arie Arnon, Alain de Janvry, Claude Fischer, Botond Koszegi, Ethan Ligon, and Elisabeth Sadoulet.

I also have many people to thank for the development of each of the chapters of this dissertation. For the first chapter, Robert Anderson, Fred Finan and Avner Greif provided valuable advice and helpful comments. I am deeply indebted to Lamis Aliryani, Betty Sadoulet and Alain de Janvry and the Social Fund for Development for supporting my fieldwork in Yemen. Nada Alhajjri, Mohammed Almawri, Mervat Al-Naimee, Ahmed Barakani, Ahmed Basharahi, Marta Colburn, Roberta Contin, Khalid Moheyddeen, Kunera Moore and Tareq Yeslam provided very helpful discussions during my stay in Yemen. I also thank Garret Christensen, Juan Sebastian Lleras, Guillermo Noguera, Jonathan Rose, Anna Spurlock, Chloe Thurston, Philippe Wingender and Lanwei Yang and participants in the Berkeley Development Lunch and the Stanford Labor and Development Reading Group for helpful discussions and comments. The data collection would not have been possible without Fayadh Alnageeb's excellent research assistance and ability to coordinate my team of Yemeni researchers.

For the second chapter, I would like to first thank my father, my brother Paul and George Jakubowski for helping me collect my Freemasonry data during several visits to Iowa and many hours sitting in front of the computer. I would also like to thank Bill Krueger, the librarian at the Iowa Masonic Library, and the many staff of the Grand Lodge of Iowa that made our visits there both enjoyable and successful. Adam Kendall at the Grand Lodge of California in San Francisco was also very helpful and supportive in guiding my research. Mark Tabbert is also owed a vote of thanks for his support of my research. I would also like to thank Claude Fischer for inspiring me to study American Freemasonry. Particular thanks goes to the All-UC Group in Economic History and the Economic History Association which provided the funding that made this research possible.

And for the third co-authored chapter, we benefited greatly from the helpful comments of David Card, Stephen Cosslett, Jinyong Hahn, Michael Jansson, Patrick Kline, Richard Smith, Tom Rothenberg, and members of the Berkeley Econometrics Reading Group. We are particularly grateful to Gary Chamberlain, Guido Imbens, Justin McCrary, Geert Ridder and Enrique Sentana for detailed comments on an earlier draft. We also acknowledge feedback and suggestions from participants in seminars at the University of Pittsburgh, Ohio State University, University of Southern California, University of California - Riverside, University of California - Davis, University of Maryland, Georgetown University, Duke University, University of California - Berkeley, CEMFI (Madrid) and Pontifícia Universidade Católica do Rio de Janeiro. Portions of the current paper previously appeared in Section 4 of a NBER Working Paper titled ‘Inverse probability tilting and missing data problems’.

# Chapter 1

# Tribal Diversity, Political Patronage and the Yemeni Decentralization Experiment

## 1.1 Introduction

A vast patronage network plays a key role in the governance strategy of Ali Abdullah Saleh, the president who has led Yemen for the past 30 years. This patronage system is one of the most important forms of corruption in Yemen and, as such, has likely contributed to Yemen's weak economic growth and development.[1] Even though patronage systems are prevalent in many other developing countries, ability to study these systems has been limited by the lack of a systematic way of measuring patronage.[2] In this paper I study how traditional elites can affect this important form of corruption and whether decentralization reforms are an effective way of reducing it.

Tribes are a traditional elite group that have been co-opted into the modern Yemeni state by a patronage network with two key components.[3] The first component of that network is formal, though secretive, with the Ministry of Tribal Affairs paying salaries, both in cash and in kind, to nearly 6,000 tribal sheikhs throughout the country (Phillips 2008). The second component functions through the schools and health facilities and provides both "ghost" employee contracts, in which a salary is paid to a non-working employee, as well as direct cash transfers (ARD 2006).[4] Patronage in the education sector alone accounts for over 6% of the entire Yemeni government budget.[5]

In this paper I develop an empirical approach for studying the role of the education sector in Yemen's tribal patronage network. First, I create a dataset of unique tribal affiliations in nearly 700 administrative units in northern Yemen. Second, I estimate two measures of the prevalence of patronage: the number of excess pupils and the number of excess teachers. These are calculated as the difference between the values reported by the Ministry of Education in educational surveys and estimates of the true values available in the population census.

[1]A variety of authors have discussed how corruption can contribute to weak economic growth and development (cf. Shleifer and Vishny 1993, Mauro 1995, Bardhan 1997, Meon and Sekkat 2005)

[2]See Odhiambo, Mittulah, and Akivaga (2005) and Kodi (2007) for discussions in Kenya and the DRC, respectively.

[3]The resulting political system has been called a 'tribal republic' because of the important role that tribal alliances play in sustaining a relatively weak central government (DRI 2008). The political system has also been referred to as 'pluralistic authoritarianism' by Posusney and Angrist (2005) who suggest that the president and a few key tribal leaders have overwhelming power.

[4]"Ghost" employees are particularly common in the school and health sectors throughout the developing world. See Reinikka and Svennson (2006) for a review. Robinson and Verdier (2003) indicate that employment is an ideal method for the distribution of patronage.

[5]Based on the fact that school employee salaries account for approximately 15% of the total national budget and 40% of school employees are believed to be ghost employees, ghost employees alone account for approximately 6% of the total budget (ARD 2006, World Bank 2006).

My dataset of tribal affiliations provides a tool for measuring tribal diversity. Drawing on the fact that neighboring tribes in Yemen are typically antagonistic towards one another, tribal diversity is measured as the number of unique tribes in a given area (Dresch 1986, Dresch 1989; Weir 2007).

I use this dataset to construct measures of tribal diversity. These measures are constructed for districts and sub-districts, the two smallest of the three basic administrative structures in Yemen. Sub-district tribal diversity is calculated as the number of tribes within a sub-district, which contain an average of 13 villages, 6,500 residents and 5 tribes. As governmental funding is distributed by district-level offices, this represents the impact of tribal diversity on the ability of a cluster of villages to attract patronage. District tribal diversity is calculated as the number of tribal confederations – a supra-tribal structure containing 2 or more tribes – within a district, which contain an average of 100 villages, 50,000 residents and 9 tribal confederations. Since tribal confederations play an important role in politics at the governorate-level, the third and largest administrative structure, this measures the impact of tribal diversity on the ability of a district to extract patronage from the governorate.

Access to educational facility data and population censuses provides a tool for measuring patronage. A first measure is the number of ghost pupils in each village. Villages receive funding for new schools, materials and teachers based on the official number of pupils and thus benefit financially when this number is exaggerated. The number of ghost, or excess, pupils is calculated as the difference between the number of pupils reported in the educational facility data and an estimate of the true number of pupils available in the population census.

A second measure of patronage is the number of ghost teachers. I estimate the number of actual teachers in a village as the product of the true number of pupils, from the population census, and the governorate teacher-pupil ratio which is calculated from the educational survey data. The number of ghost teachers is then the difference between the reported number of teachers and this estimate of the probable number of teachers.

My empirical strategy is based on the two steps of the Ministry of Education's funding mechanism. In the first step, district-level authorities prepare budgets of educational resources needed for each village. In order to study the impact of tribal diversity on this budgeting process, I create sub-district aggregates of each measure of patronage. I then estimate the impact of sub-district tribal diversity on these aggregates using a regression that includes district fixed effects and economic, population and topographic controls.

In the second step of this funding process, Ministry of Education offices in each governorate review district budgets and allocate funding to each district. Using the same sub-district aggregates of patronage, I estimate a regression including governorate fixed effects, district and sub-district tribal diversity, and a variety of economic, population and topographic controls. The point estimate on district-level diversity measures the impact of this diversity on the aggregate patronage obtained by that district.

I find that district-level diversity has a significant negative impact on both measures of patronage. Thus, districts with more tribal confederations have fewer ghost pupils and ghost teachers. This suggests that districts with more tribes face difficulties in forming unified coalitions to extract patronage from the state.

I also find evidence that sub-district diversity increases the quantity of ghost pupils and ghost teachers within districts. The contrast between this result and the previous result suggests that tribes play a different role in district and governorate politics. Most tribes have limited power in governorate-level politics. However, each tribe within a district has significant influence as they can make credible threats to disrupt local economic activity.[6] District authorities will thus allocate relatively equal quantities of patronage to each tribe to minimize the chances of conflict and sub-districts with more tribes will receive a larger share of the total patronage allocated by a district.[7]

Throughout the analysis, identification rests on the assumption that the number of tribes and tribal confederations is exogenous to the patronage system. The validity of this assumption is supported by the fact that the educational patronage system developed very recently. Since the first rural schools did not appear until the 1960s, and very few rural areas had schools before the mid-1970s, it is unlikely that the educational system affected the borders of the of sub-districts that were demarcated during the 1930s-1960s and have remained largely stable since then (Phillips 2008). Similarly, the tribal order has been quite stable for at least 1,000 years, despite repeated Ottoman occupations, the rule of several Imams and the arrival of the Republic. It is unlikely therefore that the education system has had a significant impact on its structure (Dresch 1989, Weir 2007).

That no unobserved factor is driving both the number of tribes as well as educational outcomes is impossible to fully verify. However, the empirical analysis contains most variables that could plausibly influence both outcomes, e.g. controls for population density, total population, number of villages, government services, agricultural conditions, and ruggedness of the terrain.

[6] An important and common example of this is the blockading of roads which is a key strategy used by tribes. As there is typically only one main road in a district and it travels through most sub-districts, any tribe along the route could effectively blockade the road (at least for some time). The reason that this is not as valuable a technique in governorate-level politics is two-fold. First, the larger roads that would be required for this type of threat travel through only a few districts and are generally well protected. Second, governorate-level conflict is very costly for all parties involved and only used as a last resort.

[7] The distribution of patronage is also likely observable locally, within districts, as they districts are quite small and there is a significant amount of participation of the tribes in district-level government offices.

An important change in the institutional framework in my sample period was a decentralization reform. This reform created popularly elected local councils designed to increase local oversight of centrally-funded development activities. It had been advocated by the government and foreign advisors as a way of reducing corruption and weakening the elite control of the distribution of development goods.[8] A significant concern is that these reforms benefited only tribal elites, as has been found in other developing country contexts.[9] Instead of giving power to local populations, most observers have indeed indicated that the elections accompanying the Yemeni decentralization reform served to empower the local, and typically tribal, elite (Boase 2001, DRI 2008, Phillips 2008). These local elites were, in effect, given the tools to engineer their desired electoral outcomes as they designed the boundaries of the electoral constituencies, had the exclusive right to disqualify potential candidates, and were actively involved in both voter fraud and intimidation without any risk of punishment (Spinelli 2003, IFES 2005).

I examine the impact of this reform on each of the two stages of the budgeting process. Since educational survey data is available for both before and after the reform, I pool all years of the data and augment my regression to include an interaction term between the year of the survey and the measures of tribal diversity.

I find no evidence that these reforms either attenuated or enhanced the influence of tribal elites. In particular, the relationship between tribal diversity and patronage does not change after the reform. I do find that the total quantity of patronage increased after the reform, though this may simply represent a secular increase in corruption in the education sector rather than a direct impact of the reform.

My analysis offers two other insights about the role of social institutions in development. First, it suggests that the ability of decentralization reforms to reduce local corruption and improve efficiency is limited. This is an important finding in the context of Yemen, as prominent Western scholars have called for continued decentralization arguing that the 2001 reform did not give local councils sufficient autonomy or control over resource distribution (Romeo and El Mensi 2008).[10] Decentralization reforms have also been proposed in many countries that

[8]See "Control of Corruption and Transparency", Yemen News Agency (SabaNet), Abdul-Salam Al-Korary, September 11, 2005 for a discussion (downloaded from `http://www.sabanews.net/en/news63365.htm` on November 2, 2009). Boase (2001) discusses how these reforms could be a force for reducing local corruption, though he does also indicate that they could provide an opportunity for increased corruption.

[9]Crook (2003) discuss elite capture in Africa and Véron, Williams, Corbridge, and Srivastava (2006) provide a similar discussion for India. See Bardhan and Mookherjee (2005) for an overview of elite capture, corruption and decentralization.

[10]Admittedly, these authors discuss other possible benefits of decentralization in Yemen including improved state stability. However, there is also little empirical support offered for this

are similar to Yemen, such as Afghanistan and Somalia.

My analysis also shows the important role that pre-Islamic institutions can play in the modern development of countries of the Muslim Middle East. Most discussions of the role of institutions in development in the Middle East have focused on Islam (cf. Kuran 2004, Chaney 2008). However, Yemen had a developed tribal society before the arrival of Islam, as did many other Middle Eastern countries (Khoury and Kostiner 1990).[11] Additionally, while pre-Islamic institutions may have a direct impact on development, as I have shown here, they may also have indirect impacts.[12] Thus, a more careful examination of pre-Islamic institutions is important in understanding the effects of Islam on development.

In the following section I discuss the role and structure of tribes in the recent history of Yemen. Section 1.3 then discusses the empirical approach including a motivation and description of the empirical estimation, the identification strategy and the measure of heterogeneity employed. Section 1.4 describes the data. Section 1.5 discusses the structure of the education system and its role in the patronage system and then Section 1.6 analyzes the impact of tribal diversity on this component of the patronage system. Section 1.7 describes the decentralization reforms and presents my evidence that suggests that elite capture accompanied this process. Section 1.8 concludes.

## 1.2 The Tribes of Northern Yemen

As a vast literature has demonstrated, the tribes, bands of individuals who group together on the basis of territorially-based identities, are the central social organization in the mountains of rural Yemen (Dresch 1989, Weir 2007). The tribes of Yemen, which emerged to help communities diversify agricultural risk and provide economies of scale in defense, have existed in a very similar form for over a thousand years.[13] In the absence of a state that could effectively reach the rural, mountainous regions of Yemen, villages banded together to form small tribal states and elected leaders that governed in both local and external affairs. They continue to perform the same role today. They have recognized leaders, a functional legal system and a set of institutions designed to prevent internal and external conflict. And with semi-democratic elections for leaders, the ability to tax local

possible effect.

[11] In the apocryphal tale describing the arrival of Islam to Yemen, the Islamic cleric that eventually started Islam in Yemen came originally at the behest of two powerful tribes in order to settle a dispute.

[12] An important indirect impact is the effect that these pre-Islamic tribal institutions have had on the development of legal systems across the Middle East (cf. Charrad 2001, Weir 2007).

[13] The early roles of tribes is hinted at in the work of Weir 2007. This will be the focus of my following project about the Yemeni tribes.

populations and a near exclusive power over the use of military force within that tribe's territory, each tribe functions as its own nation-state.[14]

In the next four sub-sections I highlight the four aspects of the tribes which are key for the current analysis. The first motivates the idea of tribal diversity by describing how tribal identities are created and maintained. The second discusses the relationship of the tribes to the state with a particular focus on the tribal patronage system. The third discusses the stability of the tribes in Yemeni history, which is important for my identification strategy discussed in Section 1.3, and the fourth discusses the functionality of the tribes and the tribal confederations which are the two main types of tribal structures used in my analysis.

### 1.2.1 Tribal Diversity

Tribal identities in Yemen are constructed through mutual opposition with other tribes (Dresch 1986, Dresch 1989, Weir 2007). And while these types of identities are often conceptualized as bonds of kinship, in Yemen they are geographical and territorial. The implication of these observations is that political cooperation will be more difficult in areas with more tribes as their mutual opposition will prevent the easy formation of political alliances.

The mountains of northern Yemen are divided, politically, like a chessboard in that each tribe is a political unit that has hostile relations with its neighbors but is allied with its neighbors' neighbors.[15] And while each tribesman identifies with his own tribe, this identity only has meaning in opposition to the hostile tribes that are his neighbors (Dresch 1986, Dresch 1989). Importantly, however, is that these opposing tribes will often unify together if they are challenged by a tribe from a distant region that is not aligned with either of them.[16] This structure has often been referred to as the "segmentary model" as any individual tribe (segment) "sees itself as an independent unit in relation to another segment of the same section, but sees both segments as a unity in relation to another section"

[14] I am not the first to suggest that tribes be conceptualized as states. Weir (2007) proposed exactly that in her rejection of the segmentary lineage system (discussed below). While she does agree that the tribes do have many characteristics of a segmentary system - e.g. nested groups, political relations expressed in kinship groups, collective responsibility, mediation by religious specialists - she suggests that these tribes have several institutional features that are at odds with the segmentary model. In particular, she argues that the tribes have long-lasting and stable leadership, administrative and judicial structures in addition to durable treaties and written laws with clearly specified personnel and procedures for their enforcement. From this she instead proposes that the tribes be conceptualized as "tiny sovereign domains, each governed and represented by leaders with constitutional authority and powers of office" (p. 4).

[15] This chessboard-like political arrangement is a general phenomenon throughout the Middle East as discussed by Tapper (1990). Weir (2007) discusses this phenomenon in Yemen.

[16] This unification is the basis for the tribal confederations described in Section 1.2.5 below.

(Evans-Pritchard 1940, p. 147).[17]

That tribal identity in Yemen is based on geography and not kinship is in opposition to the standard conceptualization of tribes. Tribes are typically conceptualized as "large kin groups, organized and regulated according to ties of blood or family lineage" (Khoury and Kostiner 1990, p.4). This idea is the basis for the segmentary lineage model of tribes, where tribesmen are linked genealogically in a hierarchical structure consisting of clans, tribes and tribal confederations (Gellner 1981 and Gellner 1991).[18]

However, this formulation of the tribes as a kinship network has been rejected for the Yemeni tribes by most scholars of Yemen (Stevenson 1985, Swagman 1988, Dresch 1989, Weir 2007). These authors provide a variety of evidence for this view including: (1) Marriages are predominantly either locally endogamous, in that people typically only marry from within their community, or within people from outside the tribe,[19] (2) communities can switch allegiance from one tribe to another, though this happens very rarely, and (3) members of these tribes cannot substantiate a genealogical relationship to the tribe when asked.[20]

### 1.2.2 Politics and the Tribes

> "The state is part of the tribes, and our Yemeni people is a collection of tribes."
>
> -President Ali Abdullah Saleh[21]

The President of the Yemen Arab Republic (North Yemen), provided this now famous response when the following question was posed to him in 1986: "To what extent has Yemen succeeded in moving from the stage of tribalism to that of the state?" President Ali Abullah Saleh, unlike his two predecessors who were eliminated by tribal elements, has maintained stability in the Yemen during his more than thirty year as president by successfully coercing the tribes.

---

[17] Note that, while Dresch (1986) supports the applicability of the segmentary model to Yemen (though not the segmentary-lineage model), Weir (2007) rejects this characterization as it is rejects the notion that the tribes function as individual corporate bodies. I tend to agree with Weir (2007) and indeed the characterization of the tribe as a corporate body and state-like structure is the basis for a forthcoming paper.

[18] Many anthropologists have argued that this conceptual framework is broadly inappropriate for many of the tribes in the Middle East. See Weir (2007), p.4 and especially footnote 9 on this same page.

[19] The reason for the tribal endogamy is likely due to the Islamic inheritance law as this endogamy maintains property within the family. This idea was mentioned to me by several informants during my fieldwork but it is also mentioned by Mundy (1995) and Weir (2007) among others.

[20] Very few tribesmen can actually report more than one or two true ancestors (Weir 2007).

[21] This was part of an interview conducted in 1986 by *al-Majallah* (October 347, 1-7). The quote, as well as the question mentioned directly below, are from Dresch (1989) (p.7).

The tribes are key to Yemeni politics, and play a central role in the stability of the Yemeni state. Described aptly as a 'tribal republic', a weak central government is supported by a network of alliances between key tribes, tribal leaders and the president (DRI 2008).[22] And an important reason that these tribes stay loyal to the central government is the existence of a vast patronage network that benefits the leaders and members of tribes that cooperate with the state.

The patronage network that is used to maintain these alliances has three pieces. The first is the Department of Tribal Affairs.[23] In addition to arranging the travel plans of tribal leaders and their meetings with representatives of the central government, this ministry is responsible for paying stipends to some 4-5,000 tribal sheikhs throughout the country. These sheikhs typically receive a cash stipend between $100 and $500, though many receive much more than this, in addition to a variety of non-cash payments including vehicles, business deals or even houses (Phillips 2008).

The second is a political system that highly favors tribal leaders as popular sheikhs were, and are, encouraged to contest parliamentarian elections. In the first multi-party elections of 1993, the President made a deliberate effort to co-opt these charismatic local leaders to join his party and thus support his government which was on the eve of a civil war.[24] Since then, influential tribal leaders continue to contest elections with either the support of the President's party or the support of the lead opposition party, which is also tribal. By contesting these elections, these tribal leaders are able to secure employment for many of their followers as well as receive preferential access to development goods through their role as parliamentarians.

The third component of this patronage network is employment in a variety of government positions. While this includes jobs with active responsibilities, such as guards, it also includes a variety of "ghost" employment where an individual is paid though never expected to work. These positions, which are typically provided through either the Ministry of Education or the Ministry of Health are typically given to family members or key allies of tribal leaders (ARD 2006, Phillips 2008) . A more thorough description of Ministry of Education "ghost" employees is provided in Section 1.5.

[22]Some observers have argued that the central government has deliberately tried to revitalize the tribes as a way of governing the country (Daair 2001).

[23]Which is administered through the Ministry of Local Affairs.

[24]The value of the support of the tribes was demonstrated by their support for the President in the civil war that followed soon after these elections. Note that there is a debate in the literature on whether the tribes played an active role in supporting President Ali Abdullah Saleh during the civil war. See Dresch (1995) and Day (2001) for dissenting views.

### 1.2.3 The Stability of the Tribes in Yemeni History

Tribal stability is essential to my identification strategy as my data on tribal affiliations is recent, from 2009, while I am studying their relationship to a patronage system that has developed over the past 30. This stability is supported by the fact that the tribes of northern Yemen have borders that are well-defined, mutually recognized and change relatively rarely. Though border clashes did, and continue to, occur as tribes compete for scarce resources, there is no historical evidence that tribes would ever take territory from another tribe by force. Ambitious tribal leaders, and tribes, instead expand their power and influence by proving their political savvy and ability to negotiate to end potentially destabilizing border conflicts.

The tribes, and the borders between these tribes, are well-defined and have been for at least several hundred years, if not longer. Drawing on a chronicle of the Yemeni tribes from the beginning of the 10th century, Dresch (1986, 1989) argues that the borders of these tribes have been largely stable for a millennia or more. Weir (2007) uses internal tribal documents to similarly conclude that the tribal borders among the tribes that she studied had been stable for at least four centuries.

Though there are often disputes over borders, which are often caused by competition over resources such as grazing rights or water sources close to the borders between two tribes, there is no evidence that these conflicts affect the borders.[25] Unlike other historical contexts, where leaders and states typically expanded their influence and prestige through the use of coercion and force, the ability to negotiate and prevent conflicts is the most prized ability for a leader in the Yemeni tribal system. A successful leader typically rose to power, instead, through success in arbitration and negotiation, as he would use this ability to expand his wealth - as these leaders were paid handsomely for arbitration - and prestige. And while conflicts do turn violent, the violence is usually relatively limited and these conflicts do not affect the results of arbitration (Weir 2007).[26] Importantly, this system likely played an important role in the prolonged stability of the tribal borders.

[25]These disputes often lead to prolonged conflict between neighboring tribes that can last for upwards of 50 years NDI (2007).

[26]The conflicting parties will continue to appeal the decision of arbitration to higher and higher tribal authorities, in a way that is not dissimilar from a Western legal system, until the cost of additional arbitration outweighs the possible benefit of a favorable verdict. The cost of arbitration increases over time as more and more influential tribal personalities, often from other tribes or confederations, are involved in the case. Note that this conclusion about the choice to conclude arbitration is my own based on my reading of legal cases discussed by a variety of authors.

### 1.2.4 Tribes

Though the initial emergence of the tribes has not been well studied, it is likely that economies of scale in defense and agricultural production played a key role. Villages are a relatively small political unit. While this allows the maintenance of stability with relatively primitive institutions, the village unit is too small for the development of cost-effective defense and is also too small to allow the amount of economic diversification that is likely optimal. As high variance in rainfall, with the resulting water scarcity that can span an entire region, is a very strong impetus for the use of violence, the survival of a village was dependent upon linkages with a larger organization. In particular, this organization could provide both insurance against these shocks as well as protection if another neighboring village or region experience a negative shock and turned to violence.

The tribes were this organization. First, they organized local defense, which have important economies of scale, and formed a variety of defense agreements with other regional tribes.[27] Second, they developed a variety of institutions that allowed relatively sophisticated risk-sharing relationships within and across villages of the tribe. These risk-sharing arrangements extended to financial borrowing, sharing of labour during periods of cultivation, artisanal activities, investment in energy generation, informal automobile associations and, importantly, allowed for the existence of corporate bodies (Weir 2007).[28] Third, they facilitated external relations and trade by insuring members of their tribes in external affairs. For any crime or damage committed outside the tribe, the entire tribe would share responsibility and a local tax would be collected in order to make good on the resulting debt or damage.[29]

[27] See Nugent and Sanchez 1999 for a discussion of the economies of scale in defense.

[28] Though Islam does not have provisions for the existence of corporate bodies as described by Kuran (2004), tribal law in many cases did indeed develop specific provisions for their existence. Interestingly, while these economic arrangements could exist without a bureaucracy such as the tribe with all its associated costs, there is a strong bias toward forming economic links with members of the same tribe. The reason for this bias, as reported by Weir (2007), is that disputes over the arrangements are much more easily solved within the tribe where the legal structure is well understood, which suggests the important role that the tribal law plays in mediating these corporate-type bodies.

[29] Weir (2007) (p.149) reports that the "the tribe is united whatever expenses, misfortunes, deaths, or retribution might afflict it", with the only exception being if the individual committed a crime with clear intent in which case he forfeits the support of his fellow tribesmen (e.g adultery, assassinations). The importance of this institution of mutual responsibility in regulating extra-tribal affairs is highlighted by a variety of authors (e.g. Meissner 1987, p.253; Dresch 1989, p.131, 210-211; Mundy 1995).

### 1.2.5 Tribal Confederations

While the tribe is of central importance in the lives of rural Yemenis, tribal confederations play an important role in the interaction of the tribes and the central government. These tribal confederations, which are also an important part of an individual's tribal identity, are a tribal structure that emerged to link groups of tribes together and played a particularly important role in extra-tribal economic interactions. The primary role of these confederations was the protection of trade routes and markets.[30] The local tribe was held responsible for any crime committed on a road, as well as against any guests and travelers, and the punishments for either of these crimes was usually double.[31] A similar protection was given to markets, though they also served as an important source of wealth and prestige for the tribes.[32] A second role was the sharing of important resources across tribes. In particular, while most tribes maintained wilderness land for the grazing of animals and water facilities, the existence of tribal confederations as a mechanism to share these resources during times of hardship was an important institution to prevent violence between tribes.

## 1.3 Empirical Approach

In this section I elaborate my empirical approach in three sub-sections. First, I motivate and present the general empirical specification that is used for examining the impact of tribal diversity on the number of ghost employees. Second, I discuss the two measures of tribal diversity. And in the third sub-section, which discusses my identification strategy, I explain why it is appropriate to treat the tribal structure as exogenous to educational outcomes.

### 1.3.1 Estimating the Impact of Diversity on Patronage

I calculate two measures of local corruption using the available educational and population data. Both measures are based on three key variables: (1) the *reported* number of pupils, (2) the *reported* number of teachers and (3) the *true* number of pupils. The first two variables are from Ministry of Education surveys and are

[30] Weir (2007) argues emphatically that trade and the value of markets are the *raison d'être* behind the existence of intertribal relationships.

[31] The punishment for any crime or infraction committed against guest and travelers or on the road is typically doubled, as the host tribe of the guest is also insulted and this insult must be repaid. In practice, the host of the guest or the traveler is the responsible party for recovering any compensation for damages or for exacting revenge if necessary.

[32] While markets were sometimes run by a single tribe, that there were limited areas that were suitable for markets and the economies of scale probably played an important role in limiting the number of functional markets.

expected to include both real and ghost pupils and teachers. The third variable is from a population census of the Central Statistical Office (CSO) and reflects an estimate of the true value as the CSO enumerators have no incentive to inflate the numbers.

The first measure of patronage is the number of excess, or ghost, pupils. This is calculated directly as the difference between the official and true number of pupils:

$$\text{Ghost Pupils} = Pupils^{\text{reported}} - Pupils^{\text{true}} + \epsilon_R^P - \epsilon_T^P$$

where $\epsilon_R^P$ represents the measurement error of the reported number of pupils and $\epsilon_T^P$ is the measurement error in the estimate of the true number of pupils. For the main estimation, the error in the estimate of the true number of pupils is assumed to be idiosyncratic and independent of tribal diversity. However, as there may be a systematic relationship between tribes and the reporting in the population census, a robustness section considers the sensitivity of these results to this assumption. Figure 1.4 reports the distribution of excess, or ghost, pupils observed in the available data.[33]

The second measure of corruption is the number of ghost teachers. This measure requires an added assumption: governorate level offices of the Ministry of Education aim to provide an equal teacher-student ratio in all areas on average. Thus, any deviations from this average reflect local rent-seeking. I first calculate the aggregate teacher-student ratio in a governorate, $\widehat{\frac{Teachers}{Students}}$, by dividing the total number of reported teachers by the total number of reported pupils in the Ministry of Education surveys. I then use this to estimate the number of ghost teachers as

$$\text{Ghost Teachers} = Teachers^{\text{reported}} - \frac{\widehat{Teachers}}{Pupils} \cdot Pupils^{\text{true}} + \epsilon_R^T + \frac{\widehat{Teachers}}{Pupils}\epsilon_T^T$$

where $\epsilon_R^T$ is the measurement error in the reported number of teachers and $\epsilon_T^P$ is, as above, the measurement error in the true number of pupils. The same assumptions discussed for the measurement error in the estimation of ghost pupils apply here. Figure 1.5 reports the distribution of excess, or ghost, teachers observed in the three years of available data while 1.6 reports the distribution of excess teachers per 100 school age boys, another measure of the prevalence of corruption.

The baseline specification is an ordinary least squares regression. The unit of observation is the sub-district and the dependent variables considered are the aggregates of each measure of patronage. This regression includes the number of tribes within the sub-district, a variety of sub-district and district controls, and district fixed effects.

[33] All three years of Ministry of Education data are pooled together for this figure as well as the following two figures.

This basic specification, which is used for studying the within district allocations, can then be written as

$$\text{Ghost}_{i,j,k} = \alpha_{i,j} + \beta_k \; TRIBES_{i,j,k} + \gamma' Z_{i,j,k} + \epsilon_{i,j,k} \tag{1.1}$$

where $i$ indexes governorates, $j$ indexes districts and $k$ indexes sub-districts. $Ghost_{i,j,k}$ is the realization of either the number of ghost pupils or ghost teachers in sub-district $k$ (which is in district $j$ and governorate $i$), $\alpha_{i,j}$ is a district fixed effect, $TRIBES_{i,j,k}$ is the number of unique tribes within sub-district $k$, $Z_{i,j,k}$ is a vector of sub-district controls within sub-district $k$ and $\epsilon_{i,j,k}$ is a sub-district specific error term that is assumed to be i.i.d.

A second specification, which is used for studying the within governorate allocations, augments this regression to include the district-level measure of tribal diversity. This can be written as

$$\text{Ghost}_{i,j,k} = \alpha_i + \beta_j \; CONFED_{i,j} + \beta_k \; TRIBES_{i,j,k} + \delta' X_{i,j} + \gamma' Z_{i,j,k} + \epsilon'_{i,j,k} \tag{1.2}$$

where $\alpha_i$ is a governorate fixed effect, $CONFED_{i,j}$ is the number of unique tribal confederations within district $j$, $X_{i,j}$ is a vector of district-level controls within the corresponding district $j$, and the rest of the parameters are defined above. Importantly, as Equation (1.2) contains aggregate explanatory variables, i.e. district-level control variables that are simple sums of the same sub-district controls and district-level measures of the number of unique tribes, direct estimation can lead to serious biases (Moulton 1990). Thus, all the estimates reported using this second parameterization are clustered at the district level.

### 1.3.2 Tribal Diversity

Two measures of tribal diversity are used throughout this analysis. The first, tribal diversity at the sub-district level, was calculated as the number of different unique tribes in that sub-district. The second, the diversity at the district level, was calculated as the number of unique tribal confederations within a district. There are, on average, four tribes in each tribal confederation, though this number ranges from 1 to 27.

Both measures of diversity are calculated as the log of the number of unique tribes. This functional form captures the fact that the marginal impact of an additional tribe on diversity is increasing in the number of unique tribes.[34]

Figure 1.9 displays the sub-district tribal diversity data that is used through-

[34]The results are not sensitive to this functional form specification.

out this analysis.[35] In this figure, the solid lines demarcate district and governorate borders while subdistrict borders can be discerned by the changing color representing the number of tribes in the subdistrict.

This figure demonstrates the dramatic variation in the degree of tribal diversity within an administrative area. All my analysis focuses on variation within governorates, as all specifications include governorate fixed effects, so the significant variation within these areas demonstrated in Figure 1.9 (the thick black lines) is particularly important.

### 1.3.3 Identification Strategy

Identification here rests on the central assumption that the number of observed tribes, and tribal confederations, within an administrative sub-district is exogenous to the development of Yemen's patronage system during the past 30 years. This assumption implies two things. First, it implies that the patronage system has not affected the number of tribal units within a sub-district during this period. Second, it implies that there is no third, unobserved, factor that is driving both the number of tribes as well as the development of the education system. In this sub-section I will discuss the variety of evidence that supports these two important assumptions.

The first piece of this assumption, that the education system has not affected the number of tribal units within a sub-district, requires two things. First it requires that the structure of the tribal order, and in particular the number of tribal units within an area, was not affected by the development of education system. And, second, it requires that the definition of the sub-district was not affected by the education system which would also create a mechanical relationship between the tribes and education. I will address each of these in turn.

The first, that the structure of the tribal order itself was not affected by educational outcomes, is substantiated by the remarkable stability of the tribes over time. Indeed, Dresch (1986, 1989) uses historical texts to show that many tribes have maintained the same borders for over a millennium. Weir (2007), who uses legal documents maintained by the tribe themselves in her analysis, similarly concludes that the tribes of her study, which do not overlap with those of Dresch, were stable for over four hundred years. The stability of the tribes is discussed at length in Section 1.2 above. And if these tribes were indeed stable despite the repeated Ottoman occupations, the rule of several Imams and the arrival of the Republic, it is fair to conclude that these structures would be largely unaffected

[35] Note that there are some areas 'missing' in this figure. My population data does not include urban areas, so those are excluded. However, I am also missing some subdistricts in the governorates for which I have population data. I am working on filling in these missing observations.

by the very recent development of the education system which only first emerged in the 1960s and did not reach most rural areas until the 1970s at the earliest (See Appendix A.1 for more details on the recent development of the education system).

The second, that the definition of sub-districts was not affected by the educational system, follows a similar argument in that sub-district borders were demarcated before the development of the education system. Though sub-districts are a relatively recent administrative unit, their demarcation was largely completed by the mid-1970s (CPO 1974). And though there are cases of sub-district borders being changed after that, most reports indicate that these changes was almost entirely to improve their alignment with tribal borders and that they had little impact on the actual day-to-day operations of these administrative units which had already been using the mutually recognized tribal boundaries. Thus, it is unlikely that the educational system affected the definitions of these sub-districts in any significant way.

The second piece of this assumption, that there is some unobserved factor that is driving both the number of tribes as well as educational outcomes is impossible to rule out entirely. Though the empirical analysis contains many of these variables that could plausibly influence both outcomes, e.g. controls for population density, total population, number of villages, government services, agricultural conditions, ruggedness of the terrain, it is in practice impossible to include every possible variable that might influence both these outcomes. However, as including increasingly extensive controls does not seem to weaken the observed impacts of tribes, it seems unlikely that some other variable, which is certain to be highly correlated with the set of other variables included, could be driving the impact.

## 1.4 Data

In this section I describe the tribal and educational data that are at the core of this analysis. A description of the administrative structure and the variety of population, area, terrain, economic and agricultural controls use throughout the analysis is deferred to Appendix A.2 though Table 1.4 provides key summary statistics for these variables.

### 1.4.1 Tribal Data

Tribal data was compiled for a total of 692 sub-districts in 84 districts in 6 governorates.[36] Though not fully representative of Yemen, these data include nearly one-third of rural Yemen and represent a diversity of tribal-state relationships. They include: (1) The governorate of Sa'ada, which borders Saudi Arabia and where the Yemeni government is now fighting a war against a rebellious tribal group, (2) Al-Jawf, in which the central government has had little penetration because of the power of the tribes,[37] (3) Amran, the home of the central political and tribal opposition to the President, (4) Hajjah, a key transit area in the Yemen-Saudi trade and an area where recent African migrants have assimilated into the tribes,[38] (5) Al-Mahweet, whose rugged terrain is covered with some of the most spectacular terraced agriculture in the country, as well as a variety of sophisticated tribal law to govern them, and (6) Ibb, the luscious agriculture lands of central Yemen that is home to many important technocrats who have served the Republic of Yemen (e.g. prime ministers, etc.) who had originated from the tribes.

For each sub-district, I created a list of all the unique tribes and tribal confederations. These lists were created by working with a group of Yemeni research assistants native to the regions studied and collected through a combination of field visits and conversations with friends and family members.

Table 1.1 reports statistics on the number of unique tribes and tribal confederations within various administrative divisions. In particular, it reports the average and standard deviation of the number of unique units of each tribal group per sub-district, per district, per governorate and overall.

This table illustrates the great diversity of tribes within even very small areas in Yemen. Indeed, the average number of tribes per sub-district, which have an average of 6,500 residents, is nearly 5. And the average district, which has around 50,000 residents, has nearly 35 tribes and almost 10 tribal confederations. Interestingly, there is no clear relationship between the size of the sub-districts and the number of tribes (correlation coefficient = 0.01).

### 1.4.2 Education Data

The education variables that are the primary outcome of interest in this analysis are drawn from two key sources: The first is a series of educational establishment

[36] The final data set will consist of approximately 1,208 sub-districts in 150 districts in 9 governorates.

[37] Al-Jawf is unfortunately excluded from much of the current analysis as agricultural data is not available for this governorate (see Appendix A.2.3 below for more details).

[38] The observation of African migrants assimilating into the tribes is based on my own observations during fieldwork.

surveys conduced during the 1999-2000, 2000-2001 and 2005-2006 school years. The second main source is the population censuses which were conducted in 1994 and 2004. Table 1.3 provides summary statistics for all the education variables used throughout the analysis.

The three educational establishment surveys used here provide a broad range of information for the universe of schools in Yemen at the time of the survey. Most important for the current analysis, however, is that they report the official values for both the number of teachers and students for each of the schools surveyed. Though the number of teachers and students is disaggregated by gender and school level, and the analysis could in principle be disaggregated by gender and level, the current analysis focuses on the aggregate number of male pupils and male school teachers. The focus on male teachers is largely pragmatic, as female positions are rarely given as patronage, and the aggregation of the different levels is only to simplify the analysis.

In addition to the total number of teachers and the teacher to student ratio, a key variable of interest is the number of teachers relative to the total population, or in this case, the total number of potential students in an area (i.e. the population of boys). Though not observable directly in this educational survey, I can calculate this variable by merging the survey data with the population data that is available in the population census. This could be done at the sub-district level using the aggregate numbers of teachers and population. However, as the variation of the teacher to population ratio within a given region is of particular interest, as it is a measure of the equity of the allocation of teachers, I calculated this ratio at the village level. As these schools do not have a unique village code that is directly matchable to the population censuses, I matched the names of governorates, districts, subdistricts and villages using Microsoft Access (that was able to match $\sim 70\%$ of the entries) and by hand (which matched an additional 20% of the schools).

The second set of data, the 1994 and 2004 population censuses, reports the enrollment rate among boys and girls ages 6 to 15 at the village level.[39] The average and standard deviation of the enrollment rate for boys within a sub-district was calculated directly using the population of the boys in each village as a weight. The current analysis does not include girls, as they typically require female teachers which are not generally susceptible to patronage, though it should be mentioned that they do not exhibit the same strong relationship with tribal diversity observed among the boys.

A third type of educational data, educational facility censuses from 2007 and

[39]The population census data for 1994 is only available for 5 of the 6 governorates for which there is tribal data available. This is because the 1994 population census available to me does not have names and the 2001 agricultural census, which uses the 1994 village identifiers from 1994 and was used for matching, does not include Al-Jawf. See Appendix A.2.3 for more details.

2009, are also available. The first census provides data for just over 13,500 schools and the second for over 16,000. And while the latter only provides general data on each of the schools, the survey from 2007 provides relatively detailed information on the facility including the number of classrooms and the donor that funded their construction. However, while these were used to construct Figures 1.3 and 1.10 for the educational background section, they are not included in the quantitative analysis sections below as they do not offer any useful insights into either the educational patronage system or enrollment outcomes that are not available in the other surveys.

## 1.5 The Role of the Education System in Yemen's Patronage Network

This section discusses the role of the education system in Yemen's patronage network. Schools are an important part of the patronage network as they create prestige for the local community as well as employment opportunities for members of the local tribe as school administrators and teachers are typically drawn from the area of the school.[40] However, the focus of this section is the prevalence of ghost pupils and ghost teacher positions which are discussed in the first sub-section. The second sub-section then outlines the funding structure of the education system which is the basis for the empirical analysis analyzing the prevalence and distribution of these ghost teacher positions

### 1.5.1 The Role of Education in the Patronage System

The lack of any systematic monitoring and evaluation, and transparency, in the local budgeting process and its near ubiquitous presence throughout Yemen make the education system an ideal structure for distributing patronage. Patronage-type employment opportunities are probably the most important tool that the Ministry of Education uses for distributing patronage.

Overall, an estimated 40% of official Ministry of Education employees are "ghost workers", or workers that receive a paycheck though are never expected to work (ARD 2006). And many of these ghost workers are teachers (World Bank 2006).[41] Though there are no studies that examine this issue specifically, the

[40]Stookey (1974) demonstrates that tribes understood the value of school facilities as early as the 1970s in his apocryphal tale of a tribal leader who first pressured a representative of the central government to build a schools and then organized a local effort to build the school with the implicit understanding that the central government would then pay the salaries of the school's staff.

[41]There are different estimates of the prevalence of this phenomenon which was first publicized

lack of transparency in the allocation of teacher positions in rural areas, and in particular the influence that local elites and members of local councils play in creating school budgets as discussed above, suggests that these "ghost teacher" positions play an important role in the tribal patronage system.

### 1.5.2 School Financing and the Allocation of Teacher Positions

Both national and local governments are involved in the distribution of education funds. However, while the central government provides nearly the entire budget for the education system, governorates are responsible for the disbursement of over 80% of the budget.[42] Importantly, teachers' salaries are distributed almost entirely by governorate-level Ministry of Education offices though the actual allocations are decided by district-level offices.

The system for allocating teachers is a simple two-step process. First, district-level offices or representatives for the Ministry of Education prepare a budget of the teachers needed for each school in the district which is submitted to the governorate office. Second, governorate-level offices of the Ministry of Education work with the Governor to review these budgets, modify them if needed and then fund each district from the total amount of funding that was given to that governorate by the central government. That the district offices exaggerate their needs is well known though the political economy of these allocation decisions is not well studied (World Bank 2006).

## 1.6 Results: Tribal Diversity and Patronage

This section examines how tribal diversity affects the two key political processes that determine the availability and distribution of ghost pupil and ghost teacher positions. First, I will examine the allocation decision that is made within these

in the Yemen Times (see World Bank (2006) (p.22) for more details). Indeed, the estimates of World Bank (2006) of the importance of this phenomenon are significantly below those reported by (ARD 2006) though the approach used by World Bank (2006) seems to be flawed. Interestingly, I often encountered these ghost teachers as taxi drivers or elsewhere in the capital. They will often unabashedly report that they are indeed teachers in another village and that they only return home once every second week or once a month to collect a paycheck.

[42]In 2005, the Ministry of Education accounted for 17% of the total distribution of these funds which are spent primarily on construction projects and the provision of books (World Bank 2006). Interestingly, there is very significant variation in the amount of money that is allocated to each governorate. Indeed, in 2005 the governorates of former South Yemen received an average of just over 40,000 YR (US$200) per student while the governorates of former North Yemen received just under 28,000 YR (US$140) per student (World Bank 2006). Note that these estimates are not population weighted averages. and each governorate is given equal weight.

districts and the role that sub-district diversity has on the ability of a sub-district to extract patronage from the district. Second, I examine the allocation decision made by the governorate-level authorities to the districts. Here I will focus on how district-level diversity affects the total quantity of patronage that a district is able to obtain. In each case I will consider briefly the other socioeconomic factors that affect the prevalence of both the ghost pupil and ghost teacher positions, though the focus will be on the tribal diversity variables.

### 1.6.1 Diversity and Local Budgeting Decisions

I begin by focusing on the district allocation decisions. Decisions regarding the quantity of resources to be allocated to each school are typically made by a technocrat who works within the district but who is usually from somewhere else in the country. While this technocrat's responsibility is, ostensibly, to provide resources to schools based on need, these technocrats' are certainly affected by pressures from these villages as well as local and national politicians to exaggerate the needs of these schools. In this sub-section I first study how diversity and other socioeconomic factors affect this exaggeration by studying the prevalence of ghost pupils teachers within a sub-district.

Figure 1.11 provides a graphical representation of the key result. Here I plot the residuals from the regression of the number of ghost teachers against a variety of socioeconomic controls against the residuals of the same regression with the number of tribes in a sub-district as the dependent variable. This scatterplot contains all years of available data and demonstrates two things. First, it demonstrates that there is considerable variation in the number of ghost teachers. The second is that sub-district tribal diversity has a strong positive relationship with the number of ghost teachers in an area.

The positive impact of sub-district tribal diversity on the prevalence of ghost teachers within a sub-district is demonstrated in Table 1.5. Importantly, as these estimates include district fixed effects, the point estimates measures the relationship between tribal diversity and patronage within a district. Thus, the significant point estimates indicate that more diverse areas are able to obtain a larger share of the amount of total patronage received by a district.

The impact of tribal diversity on ghost teachers estimated in Table 1.5 is large. Indeed, these point estimates imply that a one standard deviation in tribal diversity would increase this measure of patronage by nearly one-fifth of a standard deviation so that tribal diversity explains nearly 20% of the variation in patronage.

Having examined the positive, significant and meaningful impact of sub-district tribal diversity on the within district allocation of patronage, I now examine the impact of other socioeconomic variables on the prevalence of this patronage. There are two interesting results that emerge from an examination of the sub-district of these controls variables reported in Table 1.6.

The first of these results is that two important classes of agriculture variables, i.e. the variables measuring agriculture wealth and the number of share of landholders with small plots, do not have a significant impact on either the number of teachers or the number of ghost teachers. The result for agriculture wealth is surprising as agricultural wealth is typically the most important component of local wealth and would be expected to affect the bargaining power of locals vis-a-vis the district-level decision making body. However, the result for land ownership is particularly striking as other authors have demonstrated for Yemen and elsewhere that can affect the prevalence of corruption.[43]

The second result is that, despite the fact that measures of agricultural wealth seem to have no impact, several of the other economic controls seem to have a significant impact on the number of ghost teachers. That only female illiteracy has a significant negative impact is interesting as this suggests that educational systems that have been dominated by men, i.e. that were educated the women of the villages when they were children, are more prone to this type of corruption.

While more educated areas seem to have more ghost teachers, the other three variables that have a significant impact suggest that ghost teachers are substitutable with other patronage-type goods and that poorer areas have fewer ghost teachers. This first result is indicated by the fact areas with less access to sanitation and electricity have more ghost teachers suggests that these are also a type of patronage and that they function as substitutes for ghost teachers. The second is demonstrated by the fact that the number of households using traditional methods for cooking – which is perhaps the best proxy for local wealth – has a negative impact on ghost teachers.

### 1.6.2 Diversity and Governorate Budgeting Decisions

I now examine the governorate-level allocation decision. Having received detailed budgets from their district-level counterparts, which they know are likely to be exaggerated, the governorate-level authority chooses a total allocation of teachers that that district will receive. While these governorate-level authorities certainly respond to the needs of the communities, pressure and coercion from both above and below affect the final allocation decision that they make. This sub-section focuses on how district-level tribal diversity affects the decision making process that determines the total allocation of teachers, and in particular, the total allocation of ghost teachers.

Table 1.7 studies the factors affecting the total allocation of patronage in

[43] See Swagman 1988 for a discussion in the Yemeni context. Banerjee, Mookherjee, Munshi, and Ray 2001 provide a more general discussion.

the available data. The key result that emerges from this table is that district-level diversity has a significant negative impact on the prevalence of both ghost pupils and ghost teachers. This result follows from the second specification, which includes sub-district diversity and is thus the most meaningful specification as observations are at the sub-district level, for each of the three variables. These point estimates imply that variation in tribal diversity can explain nearly one-third of the variation in the total number of ghost pupils and over one-third of the variation in the total number of ghost teachers.

In Table 1.8 I then examine how other district-level variables might affect the governorate-level distribution of patronage. Here I find evidence that districts with smaller populations seem to be able to secure funding for a higher concentration of teachers. This observation provides insight about the possible mechanism driving the observed impact of tribal diversity. Though the ostensible goal of Ministry of Education is to provide a roughly equal coverage of teachers within governorates, this result provides evidence that this rule is not followed. In particular, this result suggests that representation during the allocation decision by the Ministry of Education may have an important impact on the final allocation decisions. As smaller districts have an equal amount of representation as larger districts, since each district has one representative, they are able to secure a higher per student number of teachers.

This result, and in particular the limited representation that is allowed for each of the districts, indicates that the impact of tribal diversity represents weakened bargaining power vis-a-vis governorate-level authorities.

## 1.7 Yemen's Decentralization Experiment

The decentralization of control over development resources, which had been a development strategy of North Yemen in the 1960s and 1970s, was a key promise in the negotiations that led to the establishment of the Republic in 1990.[44] In 2001, after over a decade of waiting, it seemed that the promised decentralization would become a reality as elections for local councils, who would have significant budgetary control and oversight over development activities, were held for the first time. However, instead of giving power to local populations, most narrative evidence suggests that these elections served to have legitimized the local, and typically tribal, elite who were given the tools to engineer their desired electoral outcomes. In this section, in addition to reviewing the details of the reform and

[44] Decentralized control of development resources is not a new phenomenon in Yemen. During the 1970s and 1980s, Local Development Agencies played a key role in the development and maintenance of a variety of infrastructure projects (Cohen, Hébert, Lewis, and Swanson 1981). Projects were funded through a combination of remittances, *zakat*, local taxes and sometimes external funding.

the risk of elite capture suggested by others, I provide empirical evidence that the reform had no significant impact on the tribal patronage network though the volume of corruption has increased over time.

In the following sub-section I discuss the creation of these local councils and the powers that were given to these councils. A second sub-section then discusses the elite capture that seems to have been widespread in these, and subsequent, elections. The final sub-section here draws on data available from before, during and after the reform to empirically estimate the plausible impact of the reform.

### 1.7.1 The Creation of Local Councils

In 2001, after nearly a decade of discussions and promises, the Yemeni government held the first elections for local governments. The Local Authority Law, which was passed in 2000 and called for elections a year later, authorized the creation of popularly elected local councils for both districts and governorates based on the pre-existing administrative structure. These district and governorate councils were given significant control over local budget allocations and gave them the responsibility of monitoring the activities of the newly created local branches of key ministries (e.g. education, finance, etc.).

However, these local councils, which found themselves responsible for local education, sanitation, medical, taxation and security systems, were often relatively inexperienced (Boase 2001). This weakness has, arguably, left these local councils particularly susceptible to the influence of powerful local elites. And, indeed, even in areas that had competitive elections, the overall effectiveness of the elections in creating local and regional councils that increased local voice in the allocation of resources has been highly contested (DRI 2008, Phillips 2008, Romeo and El Mensi 2008).

### 1.7.2 Elite Control over the Local Council Elections

The decentralization process, from the very beginning, seemed to be designed to ensure capture by local elites. Election committees were dominated by local elites, irregularities in voter registration and the candidate nomination process were common and voter impersonation was widespread. Indeed, some observers of the election cautioned that there were indications that the decentralization and the creation of local councils was seen by tribal leaders as a way of "reinforcing their traditional authority" and led to the "institutionalization of tribes as political entities".[45] Importantly, the local capture that was endemic to these elections

[45] The possibility of this elite capture, and in particular capture by the powerful and prevalent tribal elite, was cautioned by Boase (2001) who indicated that the decentralization experiment might hasten the spread of local corruption, rather than increase accountability, and that the

seems to have persisted into the 2003 parliamentary elections and beyond as a variety of similar difficulties were faced in subsequent elections.[46] Appendix A.3 provides more details on the variety of weaknesses of this decentralization process.

### 1.7.3 Measuring the Impact of the Reform

In this sub-section I study the impact that the 2001 decentralization reform had on the relationship between tribal diversity and both the prevalence and distribution of patrongage. The approach used is relatively simple and is based on the availability of surveys of educational facilities before, during and after the reform (1999, 2001 and 2006). While there are other significant economic events happening during this period that would be likely to affect patronage, my identification relies on the fact that there are no events that would be expected to affect the local impact of tribal diversity.

In order to estimate the impact of the decentralization reform I pool the three years of data and estimate an augmented regression model that includes year fixed effects interacted with diversity. I first estimate the impact of the decentralization reform on the within district allocations in Table 1.9. This is analogous to the approach considered in Section 1.6.1 and regresses the number of ghost teachers and ghost pupils against sub-district diversity interacted with the year of the survey as well as a variety of controls.

Table 1.9 demonstrates that this reform did not weaken the ability of tribes to secure patronage. Indeed, I actually find weak evidence that the prevalence of ghost pupils and ghost teachers actually increased in the wake of the reform. This is seen by the weakly significant point estimate of the diversity interacted with the 2006 year dummy. However, the point estimate on the 2001 interaction term suggests that this may simply reflect a secular change in the influence of the tribes rather than a direct impact of the reform.

In Table 1.10 I use the same approach to examine the affect of the reforms on the governorate-level allocation process. If decentralization did improve local voice, I would expect that the reform would attenuate the negative impact of district-level diversity found above. Here, I find no evidence of a significant effect

---

tribal structure might be a particular impediment to the success of the exercise. The first quote is from Phillips (2008) (p. 78) and the second is from DRI (2008) (p. 39, fn. 82).

[46] First, the approach used for delineating the electoral constituencies for the 2003 elections remained suspect with maps drawn by hand that were reviewed by local officials, with these local officals allowed to propose suggestions for changes. Second, there were still significant concerns of partisanship in the Supreme Council for Elections and Reform (SCER), the organization that succeeded the SEC, which raised concerns of partisanship in the in the disqualification of candidates in local elections (DRI 2008). Indeed, local authorities seemed able to disqualify candidates with seeming and in 2006 where over 5,000 candidates had their nominations rejected without explanation (DRI 2008, p.51, fn. 117).

as the coefficient estimates for all the interaction terms are a noisy zero.

While neither Table 1.9 nor Table 1.10 present evidence of a significant impact of the reform on the relationship between the tribes and the patronage network, they do suggest that corruption increased during this period. This is demonstrated most dramatically in Table 1.9 by the significant positive point estimates on the two year variables. However, though the point estimate for 2006 is significantly larger than that in 2001, that the point estimates for both 2001 and 2006 are large and significant is evidence of a secular increase in the amount of corruption rather than an impact of the reform.

## 1.8 Conclusion

This paper has made two contributions to our understanding of the functioning of patronage networks in the developing world. The first one is to demonstrate, using a new dataset of unique tribes that covers nearly one-third of rural Yemen, that a type of diversity that is neither religious, ethnic nor political can have a significant impact on the functioning of a patronage network. This diversity affects both the average level of patronage received by a region as more diverse areas have less patronage employment opportunities.

These findings demonstrate that a pre-Islamic institution can have an important impact on development outcomes in a country of the Muslim Middle East. Current discussions of development in the Middle East are dominated by religion with little recognition of the diversity of historical experiences and the role that they might play in development. The inefficiency that can be caused by a non-Islamic institution, as development resources are diverted to rent-seeking tribal groups, underscores how an improved understanding of these structures is important for understanding development in the Middle East.

Second, my findings illuminate the patronage system that is key to the stability of Yemen's 'tribal republic'. While other observers have commented on the importance of the tribes in this patronage system, mine is the first to document the relationship between the tribes and the state. This relationship is of perhaps particular relevance given Yemen's stability during the last 30 years, despite its political, historical and geographical similarities to Afghanistan and Somalia. This suggests that embedding tribal elements in a patronage system may dominate the divide and rule approach to handling 'warlords' that are more commonly employed in these countries.

My second contribution is to demonstrate how a decentralization reform can enhance local corruption. While local tribal diversity did not have a significant relationship with the distribution of patronage before the reform, I find evidence that this effect increased after the election. Thus, a reform that was designed to weaken the power of the entrenched tribal elite seems to have done just the opposite. Understanding the implications of this elite capture is of particular importance as decentralization reforms are still being promoted as the panacea of modern development, in Yemen and other countries with a similarly powerful traditional elite.

Figure 1.1: Illiteracy Among Men Ages 10 and Over in 1994 and 2004

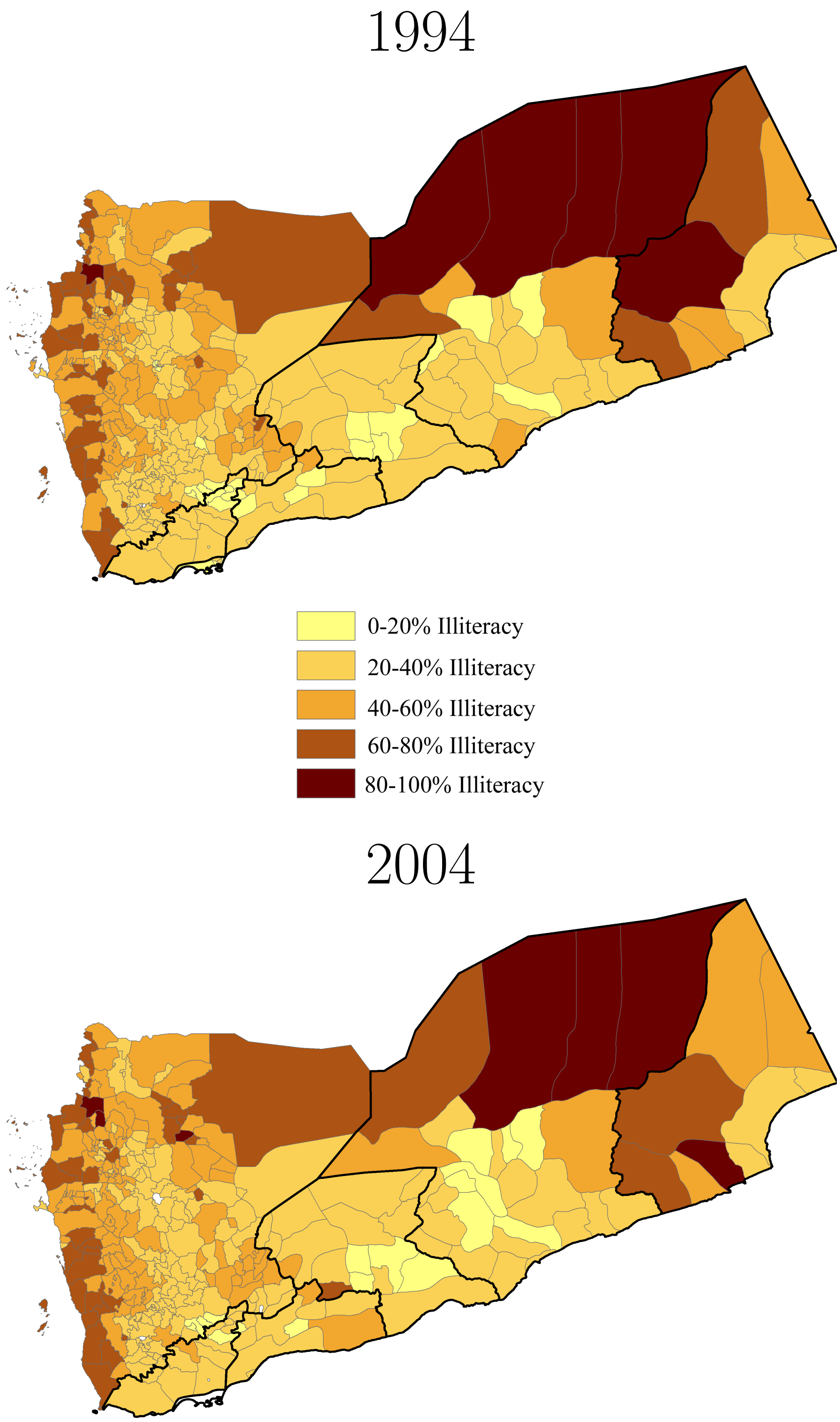


Figure 1.2: Male Enrollment Rates in 2004

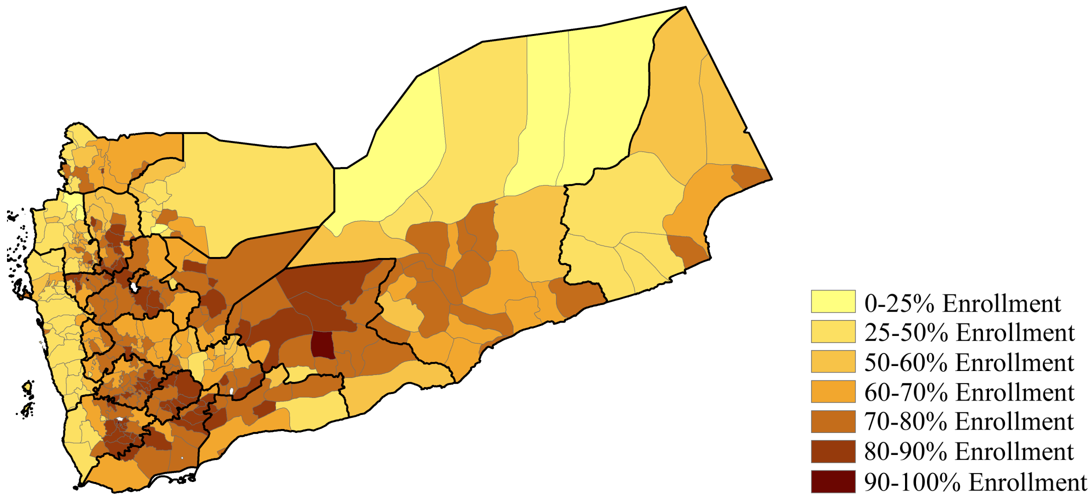



Figure 1.3: School Construction in the Northern Governorates

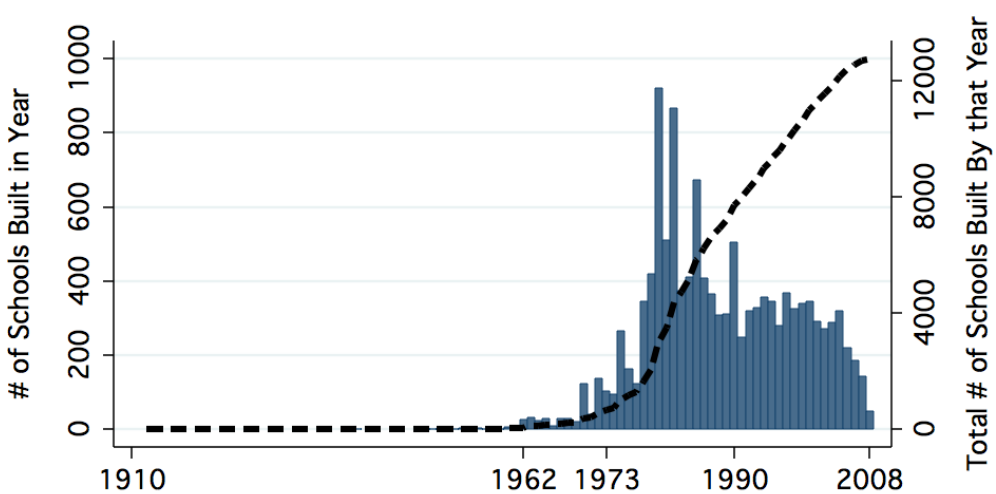


Figure 1.4: Distribution of Number of Ghost Pupils (per Subdistrict, 100s)

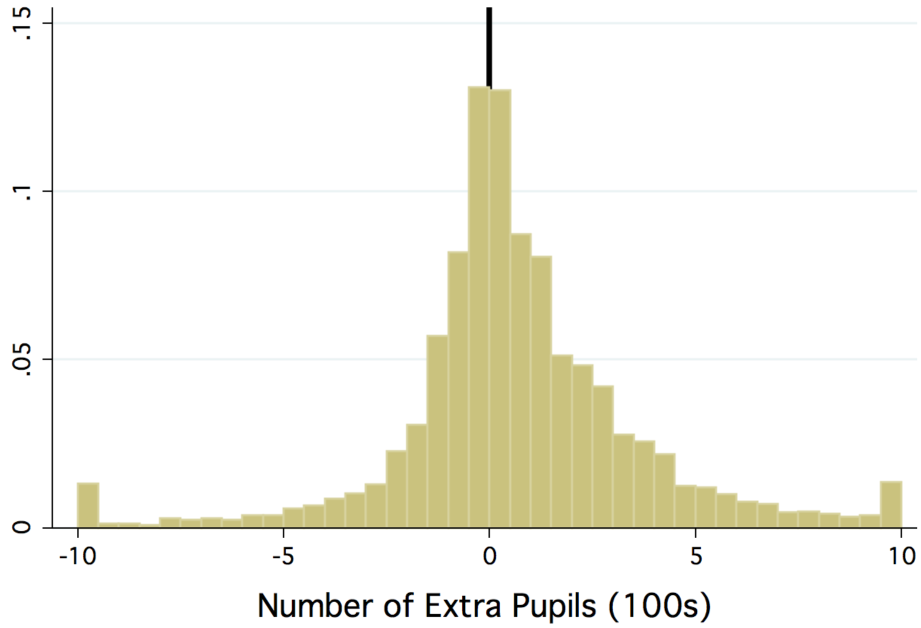



Figure 1.5: Distribution of Number of Ghost Teachers (per subdistrict)

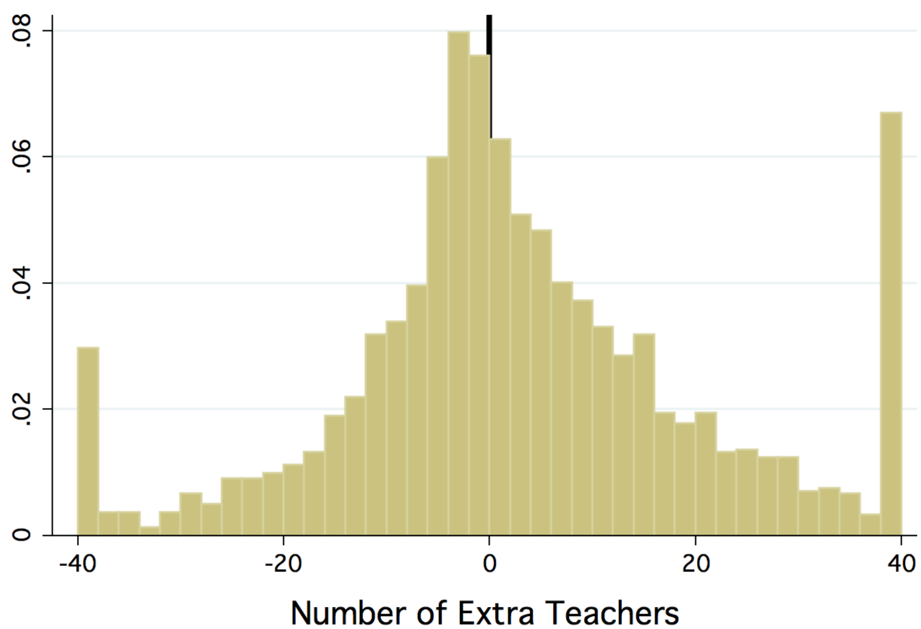


Figure 1.6: Distribution of Number of Ghost Teachers per Potential Student (per subdistrict)

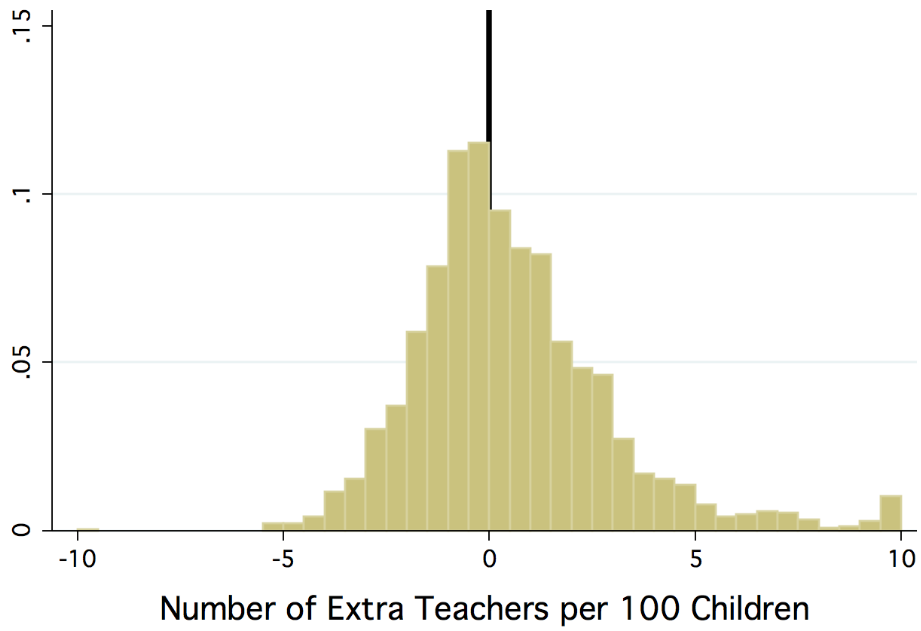



Figure 1.7: Tribal diversity in Al-Jawf and Saada

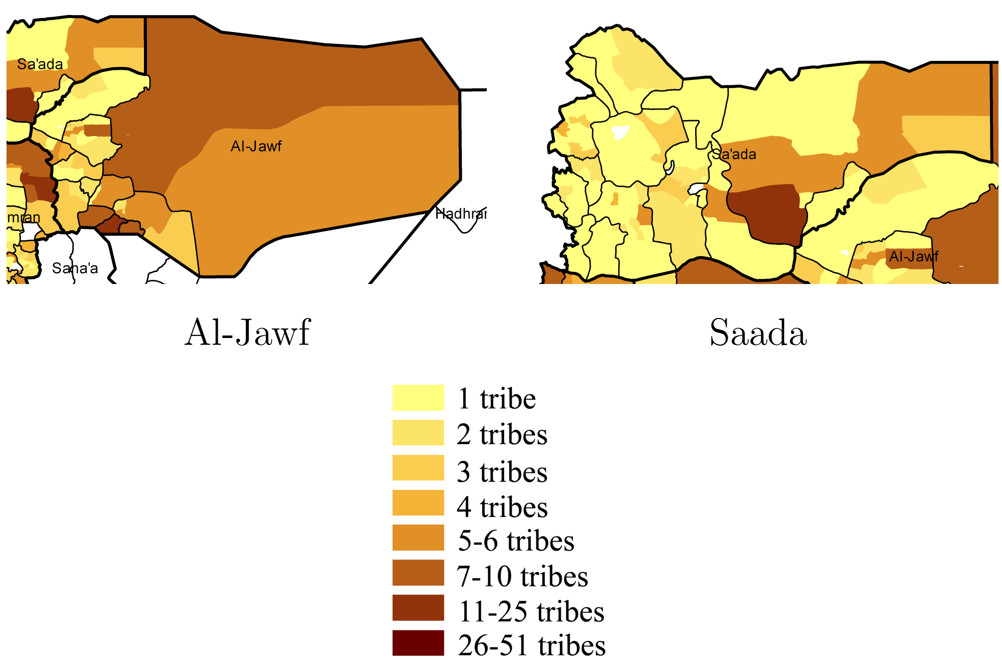


Figure 1.8: Tribal diversity in Amran and Al-Mahweet

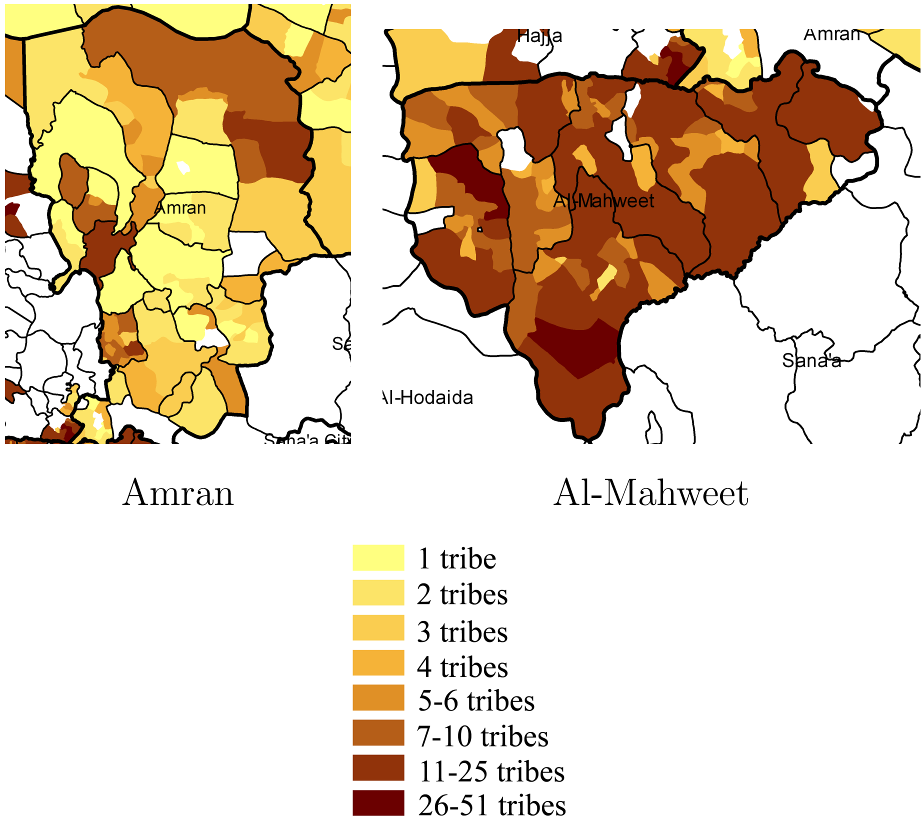


Figure 1.9: Tribal diversity in Hajjah and Ibb

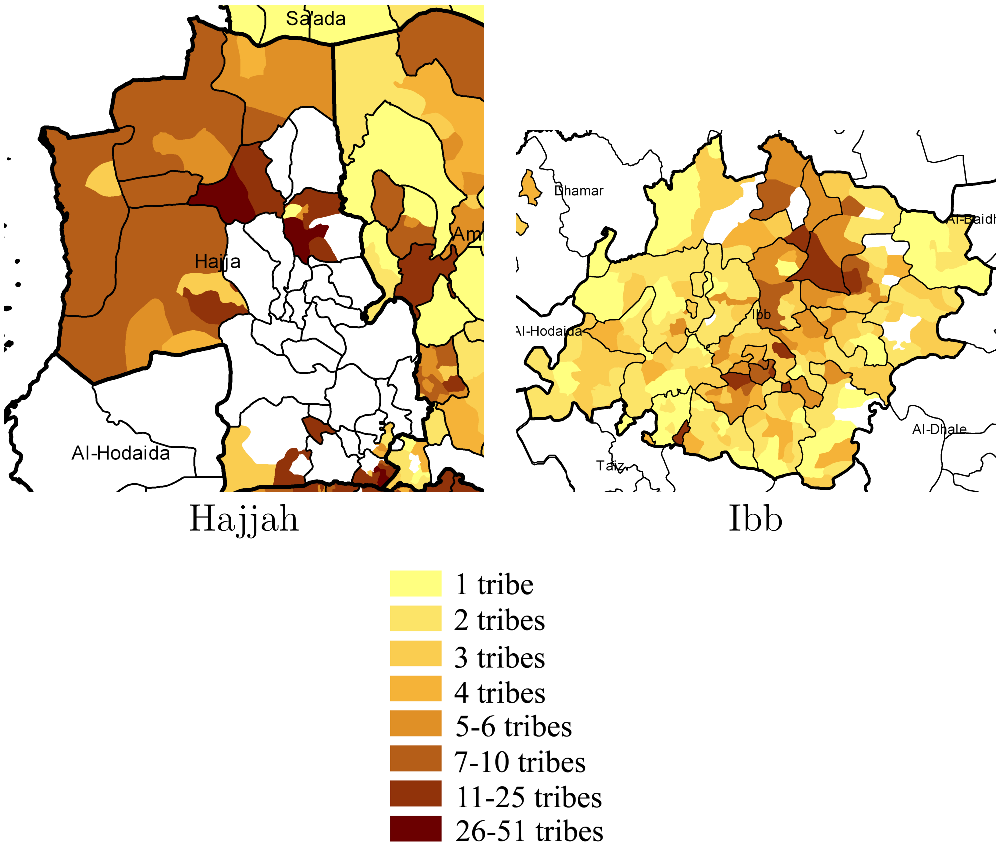



Figure 1.10: School Room Construction in the Northern Governorates

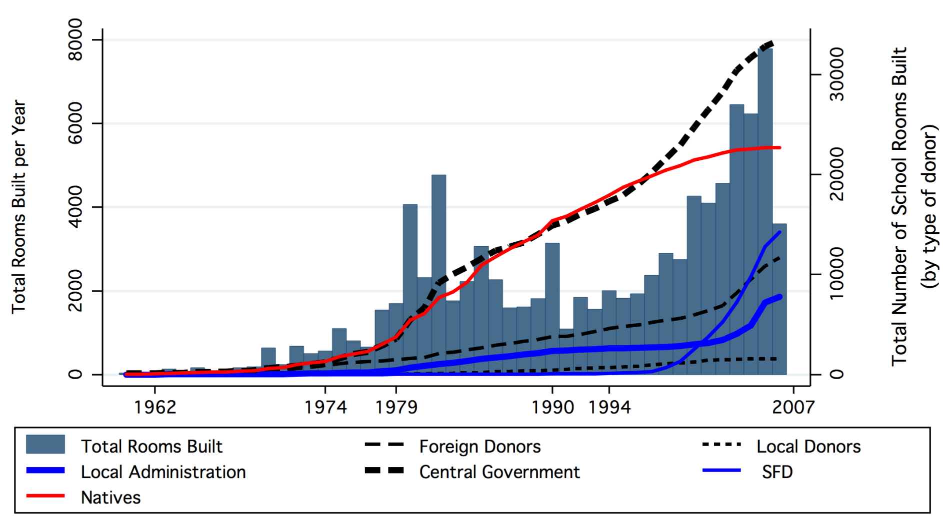


Figure 1.11: Scatter Plot of Ghost Teachers and Number of Tribes

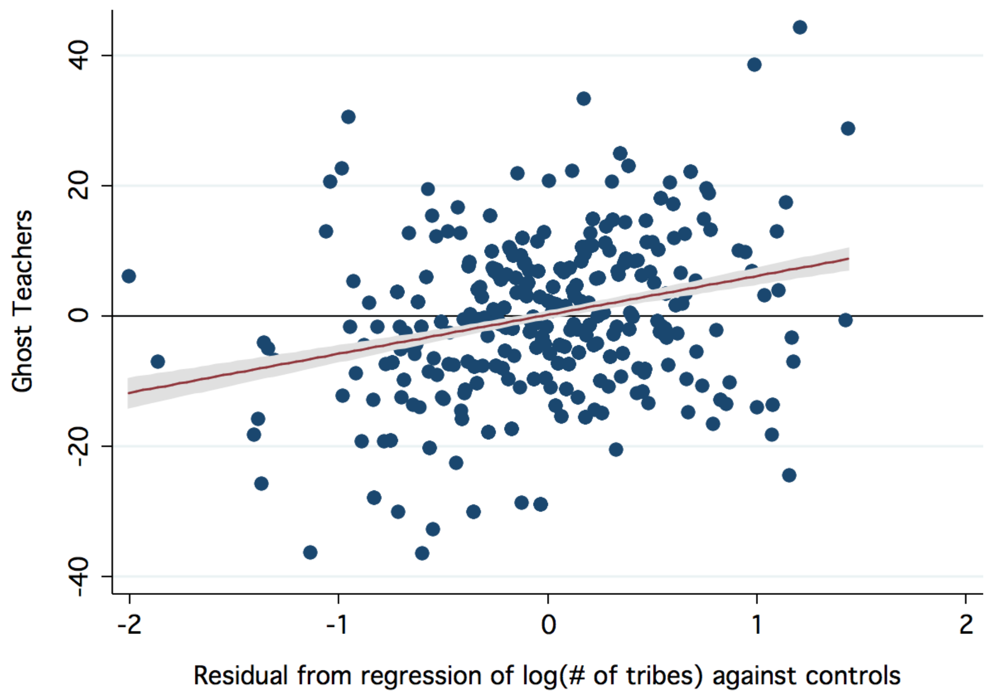


Table 1.1: Summary Statistics for Tribal Data

| | Mean Number of Unique Tribes per… | | | |
|---|---|---|---|---|
| | **Subdistrict** | **District** | **Governorate** | **Overall** |
| # of Confederations | 1.76 | 9.42 | 130.4 | 645 |
| | (1.11) | (11.51) | (100.9) | |
| # of Tribes | 4.69 | 34.38 | 426.4 | 2466 |
| | (5.32) | (45.2) | (383.7) | |
| N = | 679 | 84 | 6 | 1 |
| Number of Villages = | 10,993 | | | |
| Number of Residents = | 5,254,153 | | | |

Table 1.2: Measures of Tribal diversity

| | **Mean** | **Standard Deviation** | **Observations** |
|---|---|---|---|
| log(# of tribes within SUB-DISTRICT) | 1.06 | 0.93 | 674 (sub-districts) |
| log(# of tribal confederations within DISTRICT) | 1.92 | 1.17 | 84 (districts) |

Table 1.3: Summary Statistics for Key Educational Variables

| Variable | Unit | Year | Mean | Standard Deviation | Observations |
|---|---|---|---|---|---|
| Reported Number of Male Teachers | Total | 1999 | 37.7 | 38.0 | 719 (sub-districts) |
| | | 2001 | 43.4 | 44.3 | 719 (sub-districts) |
| | | 2006 | 47.4 | 47.6 | 725 (sub-districts) |
| Reported Number of Male Students (100s) | Total | 1999 | 6.6 | 6.8 | 719 (sub-districts) |
| | | 2001 | 7.4 | 7.3 | 719 (sub-districts) |
| | | 2006 | 8.1 | 8.1 | 725 (sub-districts) |
| Number of School Age Boys (100s) | Total | 1999 | 9.5 | 11.2 | 687 (sub-districts) |
| | | 2001 | 9.6 | 8.8 | 687 (sub-districts) |
| Number of Boys in School (100s) | Total | 1994 | 6.6 | 9.5 | 687 (sub-districts) |
| | | 2004 | 6.6 | 5.9 | 687 (sub-districts) |
| Estimated Number of Ghost Pupils (100s) | Total | 1999 | 0.2 | 4.0 | 672 (sub-districts) |
| | | 2001 | 0.9 | 3.6 | 672 (sub-districts) |
| | | 2006 | 1.6 | 3.8 | 677 (sub-districts) |
| Estimated Number of Ghost Teachers | Total | 1999 | 0.3 | 23.9 | 672 (sub-districts) |
| | | 2001 | 4.8 | 23.5 | 672 (sub-districts) |
| | | 2006 | 8.8 | 25.1 | 677 (sub-districts) |
| Estimated Number of Ghost Teachers per 100 Potential Pupils | Total | 1999 | 0.1 | 2.1 | 672 (sub-districts) |
| | | 2001 | 0.5 | 2.4 | 672 (sub-districts) |
| | | 2006 | 1.1 | 3.6 | 677 (sub-districts) |

Table 1.4: Population, Area, Terrain, Economic and Agricultural Controls

| | | Mean | Standard Deviation | Observations |
|---|---|---|---|---|
| Population Controls | log(number of boys in the sub-district in 1999 [imputation]) | 6.6 | 0.80 | 674 (sub-districts) |
| | log(number of boys in the sub-district in 2001 [imputation]) | 6.6 | 0.79 | 674 (sub-districts) |
| | log(number of boys in the sub-district in 2006 [extrapolation]) | 6.6 | 0.81 | 674 (sub-districts) |
| | log(number of boys in the district in 1999 [imputation]) | 9.4 | 0.78 | 84 (districts) |
| | log(number of boys in the district in 2001 [imputation]) | 9.2 | 0.77 | 84 (districts) |
| | log(number of boys in the district in 2006 [extrapolation]) | 8.5 | 0.87 | 84 (districts) |
| | log(# villages in sub-district) | 11.9 | 10.43 | 674 (sub-districts) |
| | log(# villages in district) | 99.2 | 69.22 | 84 (districts) |
| Area and Terrain Controls | log(area of the sub-district) | 17.1 | 1.3 | 674 (sub-districts) |
| | log(area of the district) | 19.6 | 1.0 | 84 (districts) |
| | Ruggedness (VRM) in sub-district | 0.38 | 0.11 | 674 (sub-districts) |
| | Ruggedness (VRM) in district | 0.39 | 0.07 | 84 (districts) |
| Economic Controls | % Adult Males Illiterate (2004) | 44.9 | 16.0 | 674 (sub-districts) |
| | % Adult Females Illiterate (2004) | 84.5 | 11.6 | 674 (sub-districts) |
| | % of Households using Wood, Coal or Kerosense for Cooking (2004) | 40.7 | 32.9 | 674 (sub-districts) |
| | % of Households without Sanitation (2004) | 83.1 | 19.5 | 674 (sub-districts) |
| | % of Households without Piped Water (2004) | 79.8 | 27.8 | 674 (sub-districts) |
| | % of Households without Electricity (2004) | 70.1 | 35.0 | 674 (sub-districts) |
| Agricultural Controls | log(total area owned by villagers in sub-district) | 12.6 | 1.3 | 674 (sub-districts) |
| | % of Land Cultivable | 86.1 | 13.2 | 674 (sub-districts) |
| | % of Cultivable Land that is Rainfed | 62.3 | 35.2 | 674 (sub-districts) |
| | % of Cultivable Land that is Fed from Wells | 4.3 | 13.8 | 674 (sub-districts) |
| | % of households owning plots less than 5,000 square meters | 57.2 | 27.7 | 674 (sub-districts) |
| | % of households owning plots 5,000-20,000 square meters | 22.2 | 16.7 | 674 (sub-districts) |
| | Number of goats & sheep per household | 8.3 | 11.2 | 674 (sub-districts) |
| | Number of cows per household | 0.9 | 0.6 | 674 (sub-districts) |
| | % of land cultivated with grains | 51.0 | 29.6 | 674 (sub-districts) |
| | % of land cultivated with qat | 14.9 | 19.2 | 674 (sub-districts) |
| | % of land cultivated with cash crops | 3.2 | 7.8 | 674 (sub-districts) |

Table 1.5: Sub-district Tribal Diversity and Ghost Pupils and Teachers

| **Dependent Variable:** | **Ghost Pupils (100s)** | **Ghost Teachers** | **Ghost Teachers per 100 Children** |
|---|---|---|---|
| log(# of tribes in SUB-DISTRICT) | 0.43<br>(0.33) | 5.75***<br>(1.89) | 0.16*<br>(0.09) |
| Population Controls | Yes | Yes | Yes |
| Area and Terrain Controls | Yes | Yes | Yes |
| Economic Controls | Yes | Yes | Yes |
| Agricultural Controls | Yes | Yes | Yes |
| District Fixed Effects | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes |
| $R^2$ | 0.35 | 0.44 | 0.31 |
| Number of Subdistricts | 674 | 674 | 674 |
| Number of Districts | 83 | 83 | 83 |

Note: Regressions are weighted by the population of boys in each sub-district.

Table 1.6: Relationship of Sub-District Socioeconomic Variables and Ghost Pupils and Teachers

| | Dependent Variable: | Ghost Pupils (100s) | Ghost Teachers | Ghost Teachers per 100 Children |
|---|---|---|---|---|
| Area and Terrain Controls | log(number of boys in the sub-district) | -3.80** | -23.87*** | -1.05*** |
| | | (1.80) | (9.06) | (0.23) |
| | log(# villages in sub-district) | 0.12*** | 0.86*** | 0.03*** |
| | | (0.03) | (0.16) | (0.01) |
| | log(area of the sub-district) | 1.51** | 9.12*** | 0.38*** |
| | | (0.59) | (3.07) | (0.12) |
| | Ruggedness (VRM) in sub-district | 2.00 | 26.37*** | 1.02** |
| | | (1.58) | (9.61) | (0.51) |
| Economic Controls | % Adult Males Illiterate | 0.12*** | 0.59*** | 0.01* |
| | | (0.04) | (0.19) | (0.01) |
| | % Adult Females Illiterate | -0.23** | -2.16*** | -0.08*** |
| | | (0.10) | (0.53) | (0.01) |
| | % of Households using Wood, Coal or Kerosense for Cooking | -0.04*** | -0.28*** | -0.01*** |
| | | (0.01) | (0.06) | (0.00) |
| | % of Households without Sanitation | -0.08*** | -0.21 | 0.00 |
| | | (0.03) | (0.15) | (0.00) |
| | % of Households without Piped Water | -0.01 | -0.08 | -0.00* |
| | | (0.01) | (0.05) | (0.00) |
| | % of Households without Electricity | 0.05** | 0.39*** | 0.01*** |
| | | (0.02) | (0.12) | (0.00) |
| Agricultural Controls | log(total area owned by villagers in sub-district) | -1.25 | -6.33 | 0.02 |
| | | (0.77) | (4.00) | (0.10) |
| | % of Land Cultivable | 0.02 | 0.10 | 0.00 |
| | | (0.02) | (0.10) | (0.01) |
| | % of Cultivable Land that is Rainfed | 0.03*** | 0.21*** | 0.01** |
| | | (0.01) | (0.05) | (0.00) |
| | % of Cultivable Land that is Fed from Wells | 0.05** | 0.10 | 0.00 |
| | | (0.02) | (0.09) | (0.00) |
| | % of households owning plots less than 5,000 square meters | -0.02 | -0.20 | 0.00 |
| | | (0.03) | (0.17) | (0.01) |
| | % of households owning plots 5,000-20,000 square meters | 0.05* | 0.21 | -0.00 |
| | | (0.03) | (0.15) | (0.01) |
| | Number of goats & sheep per household | -0.10* | -0.42 | 0.01 |
| | | (0.05) | (0.27) | (0.01) |
| | Number of cows per household | 2.94** | 11.78* | -0.12 |
| | | (1.24) | (6.22) | (0.13) |
| | % of land cultivated with grains | -0.00 | -0.01 | -0.01** |
| | | (0.02) | (0.11) | (0.00) |
| | % of land cultivated with qat | 0.03 | 0.08 | -0.00 |
| | | (0.02) | (0.13) | (0.00) |
| | % of land cultivated with cash crops | 0.04 | 0.17 | -0.00 |
| | | (0.03) | (0.18) | (0.01) |
| | District fixed effects? | Yes | Yes | Yes |
| | $R^2$ | 0.35 | 0.44 | 0.31 |
| | Number of Subdistricts | 674 | 674 | 674 |
| | Number of Districts | 83 | 83 | 83 |

Note: Regressions are weighted by the population of boys in each sub-district.

Table 1.7: District Tribal Diversity and Ghost Pupils and Teachers

| Dependent Variable: | Ghost Pupils (100s) | | Ghost Teachers | | Ghost Teachers per 100 Children | |
|---|---|---|---|---|---|---|
| log(# of tribes in SUB-DISTRICT) | | 0.62* | | 6.03*** | | 0.20** |
| | | (0.34) | | (1.95) | | (0.09) |
| log(# of tribal confederations in DISTRICT) | -0.80* | -1.11** | -4.29* | -7.32*** | -0.08 | -0.18 |
| | (0.46) | (0.52) | (2.40) | (2.75) | (0.10) | (0.12) |
| Population Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Area and Terrain Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Economic Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Agricultural Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Governorate Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.25 | 0.25 | 0.33 | 0.33 | 0.24 | 0.24 |
| Number of Subdistricts | 674 | 674 | 674 | 674 | 674 | 674 |
| Number of Districts | 83 | 83 | 83 | 83 | 83 | 83 |

Note: Includes sub-district and district population, area and terrain controls. Regressions are weighted by the population of boys in each sub-district and clustered at the district level.

Table 1.8: Relationship of District Socioeconomic Variables and Ghost Pupils and Teachers

| **Dependent Variable:** | **Ghost Pupils (100s)** | **Ghost Teachers** | **Ghost Teachers per 100 Children** |
|---|---|---|---|
| log(number of boys in the district) | -1.46** | -9.73*** | -0.70** |
| | (0.70) | (3.34) | (0.27) |
| log(# villages in district) | -0.02 | -0.13** | 0.00 |
| | (0.01) | (0.06) | (0.00) |
| log(area of the district) | 0.85** | 4.65* | -0.07 |
| | (0.42) | (2.56) | (0.12) |
| Ruggedness (VRM) in district | 2.28 | 10.14 | 0.58 |
| | (4.36) | (26.10) | (1.04) |
| Population Controls | Yes | Yes | Yes |
| Area and Terrain Controls | Yes | Yes | Yes |
| Economic Controls | Yes | Yes | Yes |
| Agricultural Controls | Yes | Yes | Yes |
| Governorate fixed effects | Yes | Yes | Yes |
| $R^2$ | 0.25 | 0.33 | 0.24 |
| Number of Subdistricts | 674 | 674 | 674 |
| Number of Districts | 83 | 83 | 83 |

Note: Regressions are weighted by the population of boys in each sub-district. Estimates are clustered at the district level.

Table 1.9: The Decentralization Reform and the Local Allocation Decision

| Dependent Variable: | Ghost Pupils (100s) | | Ghost Teachers | | Ghost Teachers per 100 Children | |
|---|---|---|---|---|---|---|
| log(# of tribes in SUB-DISTRICT) | 0.43 | 0.05 | 5.75*** | 3.02 | 0.16* | 0.15 |
| | (0.33) | (0.45) | (1.89) | (2.54) | (0.09) | (0.11) |
| log(# of tribes in SUB-DISTRICT) * (Year = 2001) | | 0.16 | | 2.23 | | -0.05 |
| | | (0.62) | | (3.30) | | (0.10) |
| log(# of tribes in SUB-DISTRICT) * (Year = 2006) | | 0.96* | | 5.86* | | 0.07 |
| | | (0.57) | | (3.15) | | (0.10) |
| Year = 2001 | 1.77** | 1.57*** | 11.45** | 8.77*** | 0.45*** | 0.51*** |
| | (0.85) | (0.59) | (4.45) | (3.01) | (0.09) | (0.14) |
| Year = 2006 | 3.42*** | 2.28*** | 19.03*** | 12.03*** | 0.90*** | 0.82*** |
| | (0.79) | (0.59) | (4.04) | (3.44) | (0.11) | (0.16) |
| Population Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Area and Terrain Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Economic Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Agricultural Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| District Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.35 | 0.35 | 0.44 | 0.44 | 0.31 | 0.31 |
| Number of Subdistricts | 674 | 674 | 674 | 674 | 674 | 674 |
| Number of Districts | 83 | 83 | 83 | 83 | 83 | 83 |

Note: Regressions are weighted by the population of boys in each sub-district.

Table 1.10: The Decentralization Reform and the Governorate Allocation Decision

| Dependent Variable: | Number of Ghost Pupils (100s) | | Number of Ghost Teachers | | Ghost Teachers per 100 Children | |
|---|---|---|---|---|---|---|
| log(# of tribes in SUB-DISTRICT) | 0.62* | 0.28 | 6.03*** | 2.15 | 0.20** | 0.11 |
| | (0.34) | (0.52) | (1.95) | (2.85) | (0.09) | (0.12) |
| log(# of tribes in SUB-DISTRICT) * (Year = 2001) | | 0.11 | | 4.75 | | 0.12 |
| | | (0.73) | | (4.08) | | (0.16) |
| log(# of tribes in SUB-DISTRICT) * (Year = 2006) | | 0.88 | | 6.76 | | 0.18 |
| | | (0.74) | | (4.34) | | (0.17) |
| log(# of tribal confederations in DISTRICT) | -1.11** | -1.27* | -7.32*** | -6.25* | -0.18 | -0.06 |
| | (0.52) | (0.68) | (2.75) | (3.51) | (0.12) | (0.14) |
| log(# of tribal confederations in DISTRICT) * (Year = 2001) | | 0.14 | | -3.11 | | -0.22 |
| | | (0.59) | | (2.97) | | (0.15) |
| log(# of tribal confederations in DISTRICT) * (Year = 2006) | | 0.34 | | 0.01 | | -0.11 |
| | | (0.59) | | (3.01) | | (0.15) |
| Year = 2001 | 1.50 | 1.08 | 9.55* | 9.77** | 0.30** | 0.59*** |
| | (0.94) | (1.04) | (5.15) | (4.83) | (0.12) | (0.22) |
| Year = 2006 | 2.24* | 0.50 | 10.96* | 2.60 | 0.27 | 0.27 |
| | (1.23) | (1.13) | (6.40) | (5.70) | (0.25) | (0.30) |
| Population Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Area and Terrain Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Economic Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Agricultural Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Governorate fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.25 | 0.25 | 0.33 | 0.34 | 0.24 | 0.24 |
| Number of Subdistricts | 674 | 674 | 674 | 674 | 674 | 674 |
| Number of Districts | 83 | 83 | 83 | 83 | 83 | 83 |

# Chapter 2

# Did Freemasonry Help Solve the Common Good Problem? An Examination of the Historical Expansion of American Education in the Western United States

## 2.1 Introduction

Public education, an essential component of development, is under-provided in most developing societies. Though a public education system benefits all, communities are typically unwilling to fund a system that is very costly, benefits only a minority of the population at any one time and provides benefits unequally (Poterba 1997). In this way, public education suffers from the classic "common good" problem.

During the 19th century, communities throughout the United States systematically overcame this common good problem facing public education. Despite strong opposition from a variety of groups who opposed public education for either financial or ideological reasons, locally based, locally funded and locally administered public education systems became widespread through this, then, developing nation (Black and Sokoloff 2006).[1] These local initiatives were the basis of the development of the most comprehensive public education system of the era, which functioned almost entirely without the involvement of the national government. Understanding the factors that allowed, and encouraged, communities throughout the United States to overcome the difficulties in the provision of universal public education has become a central strand in the historical education literature.

Previous studies have proposed a variety of mechanisms to explain the rapid spread of public education during the 19th century. Many studies have pointed to the importance of economic factors, such as the onset of industrialization or changing labor market conditions that lowered the cost of education.[2] Others have argued for the importance of religious faith, social organizations such as labor organizations or urbanization in explaining this spread.[3] However, few previous studies have considered the potential role that American Freemasonry played in the development of public education in the United States, despite the general knowledge that many of the central figures in the education movement were Freemasons and that Freemasonry was likely the most influential social organization in the United States during this period.

The provision of free primary and higher education was indeed a central tenet of 19th century American Freemasonry. Freemasonry, a voluntaristic secret society that attracted and accepted men of all backgrounds, is well known for its prominence among national politicians and businessmen. Though many of the activities of this organization are shrouded in secrecy and mystery, their support

[1] See also Cubberley (1920) and Soltow and Stevens (1981) that are mentioned by in footnote 3 of Black and Sokoloff (2006).

[2] Black and Sokoloff (2006) and Go and Lindert (2008) are two recent examples.

[3] See Boli, Ramirez, and Meyer (1985) and Black and Sokoloff (2006) as examples. Another interesting argument offered by Black and Sokoloff (2006) (p. 73), particularly in light of the region considered here, is that "residents of western states" were somehow different and that their pioneer attitude played an important role in the development of local public education.

for public education was quite open. Indeed, these men, who were committed to the American ideals of equality and opportunity for all, believed that education offered the means for individuals to obtain this equality and opportunity. This was famously argued by perhaps the most celebrated and influential Freemason of the period, Albert Pike:

> "Equality has an organ; – gratuitious and obligatory instrution. We must begin with the right to the alphabet. The primary school *obligatory* upon all; the high school offered to all. Such is the law. From the same school for all springs equal society." (Pike 1871, p.44)[4]

Here I argue that Freemasonry played an important role in supporting the expansion of education in the United States. In particular, three characteristics of 19th century American Freemasonry can explain its efforts and success in overcoming the public good problem and encouraging the development of public education. The first of these characteristics is that education, and in particular universal and free public education, was one of the most important political issues during this time for American Freemasons. The second is that the Freemason organization was particularly effective at developing relationships between men of different backgrounds, and these relationships were often stronger than those found within church organizations or even families. Third, Freemasons were particularly well suited for finding political for causes that they supported as they were very politically active and often dominated state legislatures and other political offices.

The analysis throughout this paper relies on two new sources of data. The first is data on the number of Freemasons per county for 24 states of the Western United States during the 1850-1900 period. Freemason data was collected decennially, to match the socioeconomic data available in the census, and the lodge level data was aggregated to the county-level for compatibility with the available economic data. The second source of data is a panel of counties for the Western United States that allows for the panel data analysis employed in this study. As county definitions changed frequently during this period, a panel data set was constructed by creating county aggregates which represent a geographical area that is constant over time.

In examining the relationship between Freemasonry and the development of American education I explore two possible types of impact. First, I focus on measuring the potential contemporaneous impact of Freemasonry on primary public education in the late 19th century. This period corresponds to the time in which Freemasonry's influence in society was likely most pronounced, and a time in which Freemasonry and public education were spreading rapidly throughout the United States.

[4] This quotation was drawn from Fox (1997).

Second, I examine the potential long-term impact of Freemasonry by looking at the relationship between Freemasonry and educational outcomes in the the early- to mid-20th century. Though the active influence of Freemasonry had faded by this period, this period corresponds to the rapid expansion of public secondary schools in the United States. Thus, by studying the relationship between the concentration of Freemasons in counties in the nineteenth century and educational outcomes in the early- to mid-20th century, I can examine the potential long-term impact of the institutions established by this organization.

My analysis is composed of three major sections. In the first section I examine the contemporaneous and lasting impact of Freemasonry on educational enrollment. Focusing on the county-level data educational enrollment available in the United States census, I use both OLS and fixed-effects panel data analysis to estimate the impact of Freemasonry and a variety of other socioeconomic variables on educational enrollment rates. Using cross-sectional data, I first examine the potentially changing relationship between Freemasonry and education at different stages of the development of Freemasonry and American education. In particular, I calculate simple OLS estimates of the impact of Freemasonry, social homogeneity, wealth, income and urbanization on educational enrollment. This approach, which is similar to that used by Goldin and Katz (1999) and thus allows comparability with the existing historical literature, allows me to compare the plausible impact of Freemasonry to other potentially important factors.

In order to address the concern that either unobserved heterogeneity or endogeneity are driving the results found in cross-sectional estimates, I then repeat this analysis using a fixed-effects panel data approach using a panel of educational enrollment, Freemasonry and other socioeconomic variables.[5] While standard fixed effects estimation allows me to control for unobserved heterogeneity, I also exploit the approach proposed by Anderson and Hsiao (1982) and Arellano and Bond (1991) and use lagged values of the Freemasonry variable to control for the possibility of endogeneity.

Here I find that the concentration of Freemasonry has a very strong positive association with educational outcomes in both 1870 and 1890. These cross-sectional results of the estimated impact of Freemasonry seem to be robust to corrections for unobserved heterogeneity and endogeneity. While I cannot verify the individual results from these cross-sectional results, as I need to combine the data from different years for this approach, the panel analysis produces very similar results overall.

And in what I believe is a striking example of the ‘path dependence’ of social institutions, I find that the concentration of Freemasonry in 1870 has a significant positive effect on educational outcomes 70 years later in 1940 at the close of the

[5] In theory, estimation using an instrumental variables strategy would deliver the same results. In practice, there are no clear instruments for Freemasonry as discussed in the data section below.

"second transformation" of American education. This is also evidence that the institutions that were established to provide public education in the 19th century - such as local governments with the ability to raise taxes to support education - had a very lasting impact on education in the United States.

In the second section of my analysis, I explore the second and third predictions of the mechanism of how Freemasonry affected education. Using measures of ethnic and religious heterogeneity that I construct using census data, I explore the second prediction and examine the relationship between Freemasonry and heterogeneity by augmenting my OLS and panel analysis with interaction terms between Freemasonry and heterogeneity. I then exploit data on education taxation and the share of local to other taxes again using census data to explore the final prediction.

My analysis of the relationship between Freemasonry and heterogeneity provides strong evidence that Freemasonry did attenuate the deleterious impact of heterogeneity on the provision of public education found by Goldin and Katz (1999) and others. This result, which indicates that Freemasonry helped heterogeneous communities overcome the common good problem, is found for both ethnic and religious heterogeneity and is robust to the inclusion of fixed effects and correction for endogeneity.

Analysis using county-level data on taxation - which is unfortunately only available in the 1850 and 1870 censuses - provides weak evidence of the proximate mechanism behind Freemasonry's impact on education. Though I do not find an effect of Freemasonry on taxation in the 1850 data, which is consistent with the fact that Freemasonry was still quite limited in its influence during this period, I find evidence from 1870 data that areas with more Freemasons paid more local taxes as a share of total taxes. If areas with more Freemasons were more effective in establishing local taxing authorities, as is suggested by this analysis, this would explain how the positive impact of 19th century Freemasonry could have persisted into the mid-20th century.

In a final section, I provide two types of robustness checks for this analysis. First, I examine the factors affecting the spread of Freemasonry by using the same cross-sectional and panel data techniques used in my first analytic section. This addresess the important concern that some other factor was driving the expansion of Freemasonry, and Freemasonry was simply functioning as a proxy for this factor. Second, I study the relationship between Freemasonry and education during a period in which Freemasonry was just developing in the American West and during which the influence of Freemasonry was much weaker. This allows me to test for the possibility of reverse causality in that Freemasons may be more likely to migrate to areas with more developed educational systems.

In my first robustness check I find the surprising result that the spread of Freemasonry does not seem to have been strongly affected by any of the socioeconomic variables available in the census. While I demonstrate that areas that were more heterogeneous and where a larger share of the population were of native-birth were more likely to have higher concentrations of Freemasons, no variables have a robust relationship across both the OLS and panel analysis that I consider. This, I would argue, is evidence that no other factor can explain the relationship between Freemasonry and education I observe here.

Then, in order to address the issue of reverse causality, I examine the cross-sectional relationship between Freemasonry and education in the 19th century. This period corresponds to both the early stages of the expansion of 'common school' in the nineteenth century as well as a time in which Freemasonry was finally recovering from a conspiracy that had weakened both the membership and the influence of Freemasonry throughout the country. Focusing on data from 1850, which is also the first year for which I have both Freemason membership and educational enrollment data, I find that Freemasonry has no significant impact on public school enrollment. Importantly, the analysis from this period suggests that Freemasons were not necessarily more likely to migrate or establish themselves in areas with better public schools. However it is also suggests that Freemasons were not responsible for the rapid expansion of the public primary school during the nineteenth century in the so-called "first transformation" of American education, but rather played an important role in supporting its expansion.

In the following section I discuss the two transformations of the American education and discuss the variety of literature that has studied this expansion. In Section 4 I explore the historical literature that expounds the relationship between Freemasonry and education in this country and in Section 3 I discuss the institution of Freemasonry in terms of its breadth, influence and membership in the United States. Section 5 introduces the Freemasonry data that I have collected for this project and the variety of educational and socioeconomic data used here. Section 6 examines the impact of Freemasonry on educational enrollment while Section 7 examines the mechanism behind this effect. Section 8 provides robust checks through an examination of the relationship between Freemasonry and education in 1850 and the correlates and determinants of Freemasonry membership during this period. Section 9 concludes.

## 2.2 The Rise of American Education

The American education system experienced two major transformations during the nineteenth and early twentieth century. The first was the rapid expansion of primary education throughout the United States during the nineteenth century. While the United States was already a world leader in terms of youth enrollment in 1830, with over 50 percent of youth ages 5-14 enrolled in school, by 1890 youth enrollment reached over 97 percent while the next closest county, France, had total enrollment of only 87 percent (Lindert 2004). The second transformation was the rapid rise of secondary schooling during the late nineteenth and early twentieth century.

The rapid expansion of primary education in the nineteenth century was driven by the rapid expansion of publicly supported non-sectarian schools which first occurred on a national scale in the United States between 1825 and 1850. Though this public expansion was a national movement, the funding and management of these schools during the 19th century was done primarily at a local level. The contribution from state taxes never surpassed 20 percent and state permanent funds contributed about 5 percent more. The federal government contributed nothing as efforts to establish a federal public education system were unsuccessful (Dexter 1904, Go 2008).

Though this local funding came from a variety of sources, such as fees, town taxes and private philanthropy, by the late 19th century local taxes was by far the largest source of funds for public education. Indeed, Dexter (1904) reports for the last decade of the 19th century that 68% of school funding came from local taxes, another 8-12% came from sources likely to be predominantly other local contributions, and only 15-18% of funding came from state taxes.[6] As the contribution from state taxes was only 17.3% in the mid-1870s and was lower than that before, local taxes likely played an equally important role throughout the century (Go 2008).

A number of theories have tried to explain both the timing and the geography of this aspect of American exceptionalism. Early researchers suggested that the rise of manufacturing, increasing concentrations of population and national leaders of education played an important role in this process (Cubberley 1922). Focusing on interregional differences in school enrollments within the United States, Go and Lindert (2008) similarly suggest that the affordability of schooling in northern states, because of higher labor incomes and an abundance of female teachers that lowered the price of schooling, help explain the large gap in enrollment between the North and the South. However, these authors argue that greater affordability is not a sufficient argument as it does not explain why communities would

[6]Dexter (1904) reports the following distribution of funding for education from 1890, 1897 and 1902 (p. 205). The other sources that I refer to above is the 'Other Sources' definition used by Dexter:

| | **1890** | **1897** | **1902** |
|---|---|---|---|
| Permanent Fund | 7,774,764 | 9,047,097 | 10,522,343 |
| State Taxes | 26,345,323 | 33,941,657 | 38,330,589 |
| Local Taxes | 97,222,426 | 130,317,708 | 170,779,586 |
| All Other Sources | 11,882,292 | 18,652,908 | 29,742,141 |
| Total | 143,194,803 | 191,959,370 | 249,374,659 |

be willing to pay taxes to support this common good. Instead, it was the decentralized system for funding education that was prevalent in the North, and lacking in the South, that allowed for the expansion of public education during this period.[7] This decentralized system allowed for more efficient outcomes as it liberated communities with greater demand for education to expand their schools (Lindert 2004).

In the early twentieth century, the education system in the United States experienced the rapid expansion of its secondary schools in what has been dubbed a “second transformation”. While only 9 percent of American youth had high school diplomas in 1910, more than 50 percent did so by 1940. And this expansion relative to other countries of the world was equally dramatic. The United States' lead in educational enrollment mentioned above had nearly disappeared by 1911 with the youth enrollment in Germany and France almost at par with the United States.[8]. However, by the mid-1920s the ratio of youth enrollment in France to the United States had slid to 0.70 and Germany and the United Kingdom did only slightly better with ratios of 0.78 and 0.77 respectively (Goldin and Katz 2008). However, the rise of secondary education was quite unequal in the United States and authors have focused on understanding the inter-state, inter-city and inter-county variation in enrollment and graduation rates to understand the factors that affected its expansion.

Similar to the first transformation of education in the nineteenth century, both economic factors and the decentralized nature of education in the United States have been implicated as essential to the rapid high school expansion during the early twentieth century. Focusing on variation across states, Goldin and Katz (2008) find that wealth, economic equality and social stability of communities, which they measure as the proportion of elderly, had a significant positive impact on total secondary-school graduation rates. And using both state and city-level data, these authors argue that increased ‘social distance’ within a city, proxied for with either fraction foreign born or Catholic, has a negative impact on the expansion of education.[9] And in Goldin and Katz (1999) the authors use county-level data to show that areas with a higher percentage of small communities and that were more socially homogenous, here using the share of the population in a county with native parentage, had higher rates of high school and college

[7]School districts in the South were typically much larger and state legislatures had much more influence in local governance (See Go and Lindert (2008) for references. Dexter (1904) also highlighted the difference in the structure of the school districts.)

[8]In 1911 the ratio of youth enrollment in France to that in the United States was 0.93 and for Germany it was 0.96.

[9]Interestingly they do not adopt the frequently used Herfindahl index for heterogeneity of the communities and instead focus on percent non-white, percent foreign born and percent Catholic. This would be possible using the data on religious membership available in county-level census data. See Alesina and Ferrara (2005) for a review of this approach.

attendance. Similar factors also affected the support for higher education during this period (Goldin and Katz 1998).

Thus, the existing evidence seems to suggest that the expansion of secondary schools in the United States was driven by local factors. And though economic factors do play an important role in explaining the regional variation within the United States, local social homogeneity or social capital seems to have also played an important role. Indeed, even though state and national efforts may have contributed a bit to the rise of secondary education, even the expansion of state compulsory schooling and child labor laws contributed at most 5 percent to high school enrollment (Goldin and Katz 2003).

## 2.3 Freemasonry and Education

Freemasonry has consistently supported education throughout its history. Indeed, though Freemasons are typically prohibited from discussing either politics or religion within the halls of the Lodges, their open support through both fundraising and rhetoric for the public provision of education seems to be one exception to this rule. In particular, Freemasons throughout the United States supported, both financially and politically, efforts to restrict financing for church-supported schools and encourage public spending for education (Demott 1986, Dumenil 1984).

Though there is no comprehensive examination of the role that Freemasons played in the development of education in the United States, a variety of historical studies have examined different aspects of the role that Freemasons played in the development of educational systems in different regions of the country and at different times.

The first way that Freemasons affected the expansion of public education in the United States was through the direct funding and construction of high schools, universities and other types of schools. De Witt Clinton, who was Grand Master of New York, established the New York Free School Society in 1809. This society provided free education to Freemason children with voluntary donations from Freemasons. In addition to providing free education to over 600,000 students and training 1,200 teachers before its closure in 1854, this Society served as a model for development of the public education system in New York and donated its buildings and equipment to the public school system that had been founded in 1842 (Mackey and Haywood 2003, p. 817).

Another example of the direct role that Freemasonry played in the actual construction of public schools is provided by Woods (1936). Responding to a growing divide in the access to public education between the North and the South during this period, the Masons of eleven southern states were directly responsible for the construction of some 88 educational institutions during the 1840s and

1850s.[10] 18 of these were schools were colleges which represented nearly 10% of all colleges in these states.[11] Further support of the importance of public education to the Masons is that graduates of these universities, which were either free or reasonably inexpensive, were either required or strongly encouraged to become teachers in common schools.

The second way that Freemasons supported the expansion of education was through politics. A particularly well documented example of this comes from Texas' formative years. Freemasons were highly over-represented in the first eight Texan legislatures (1846-1861). During this 16 year period, Freemasons represented over 50 percent of each legislature, every lieutenant governor and all but one governor, despite accounting for less than two percent of the total population. During this period, Freemasons introduced the idea of a permanent school fund (passed in 1856), which set aside ten percent of annual revenues for schools, and Freemason politicians were the first to introduce measures relating to free public school education in every session (Thornton 1962). In the following 25 years, from 1861-1885, Freemasons play a leading role in supporting bonds as a method of school finance, promoting and protecting school lands, and in helping develop administrative procedure for schools (Thornton 1970).

While the political influence of Freemasonry in education in other states is not as clearly documented as it is for Texas, Freemasons played leading roles in supporting public education systems throughout the 19th and in the early 20th century. The founder of the New York Free School Society, De Witt Clinton, eventually became Governor of the State of New York and has been called the "Father of Public Schools in New York" by some. During his tenure, his society was officially incorporated, received grants form the state and was eventually taken over by the state, that intended to extend this system to the requirements of the entire public (Lang 2006).

A prominent example of Freemasons in education in the twentieth century is the role played by a California Grand Master, Charles Albert Adams, who used the system of California Masonic Lodges to promulgate his establishment of Public Schools Week Observance in 1920. In the midst of a decaying public education system, over 600 schools in California had been closed due to teacher shortages and the remaining teachers were often inadequately trained, it has been argued that his efforts and the efforts of his fellow Freemasons were responsible

[10] Woods (1936) notes that the motivation for this school construction was that "public education in the South was much slower in getting under way".

[11] The state census records indicates that there were a total of 212 colleges in 1860 in the 11 states covered by Woods (1936). These 18 schools were distributed across the 11 states as follows (the first number indicates the number of colleges built by Freemasons and the second number is the total number of colleges in that state): Alabama (3/17), Arkansas (1/4), Florida (0/0), Georgia (5/32), Kentucky (3/20), Mississippi (1/13), Missouri (1/36), North Carolina (1/16), South Carolina (1/14), Tennessee (4/35), Texas (1/25).

for the successful adoption of an amendment to the California constitution that established a fixed contribution of funds for public schools.[12]

Also during this period the Masons played an important role in supporting the unsuccessful Smith-Towner Educational Bill in 1919 which advocated the establishment of a federal department of education. This bill emerged out of the same crisis in education after World War I, which had seen the closing of over 18,000 schools across the country. This department of education would have distributed funds to states to combat illiteracy, and recipient states would have to require attendance of all seven to fourteen year olds in school over at least a twenty-four month period. Masons throughout the country supported this bill, and the Supreme Council of the Southern Jurisdiction of Freemasons uncharacteristically voiced their public support for this bill, in order to provide funds for the equalization of educational opportunities for children throughout the nation (Dumenil 1990, Mackey and Haywood 2003 p. 817).[13]

A third way that Freemasons supported public education was by providing facilities for schools when specialized schools had not yet been built. Indeed, as many Freemason Lodges were two floors, so that the rituals of the organization could take place on the upper and more private level, some communities used the first floor of the local lodge as a school. Though there is suggestive evidence that this occurred throughout the West, one historian of Texas education has highlighted the important role that this played in the early development of public education in Texas:

> "The people in these pioneer communities had few facilities for public services. Even though the state might provide meager funds for paying a teacher, it provided nothing for a building, a place to teach. That the community had to furnish. But there were the Masons who had a hall, always a two-story hall because they could not hold their meetings on the first floor... In community after community this lower story of the Masonic Hall became the first house of learning, maybe a private or subscriptive school which was converted into a public school as soon as possible. This hatching of public free schools out of the nest of society could not have happened in so many cases had not the Masons been favorable to public education. And their interest was not passive - it was active. They did what was needed when it was needed." (Webb 1955)[14]

[12]Attributed to Vaughan MacCaughey, a prominent educator in the early twentieth century in California and Hawaii, in Whitsell (1950).

[13]This bill also found support from the Ku Klux Klan and the Daughters of the American Revolution who in particular believed that it would help stymy the influence of the Catholic church in education and create homogeneity throughout the nation (Dumenil 1990).

[14]This quote is drawn from Thornton (1970). The author notes that Webb was not a Mason.

Though Masons likely furnished less buildings for schools than the entire religious establishment in Texas, the services of the Masons in providing a space for public education has been acknowledged as one of the "most important transitional steps towards free public education" (Eby 1925).[15]

Finally, most lodges and grand lodges established educational funds to support the studies of sons, daughters and orphans of widows. These funds were one of the central charities provided by the Freemasons, the others generally being for the provision of the widows of Masons, and usually had quite significant amounts of money. One prominent example was the Massachusetts Masonic Education and Charity Trust which was allowed to hold up to one million dollars worth of funds when it was incorporated in 1850, though this was increased to five million dollars by 1916.[16]

Similar examples are found throughout the country. In 1847 alone, Ohio Masons called for the establishment of a $300,000 fund for the establishment of an educational facility for the youth of Masons,[17], a fund was established in Kentucky that would be endowed with some $40,000 over ten years,[18] North Carolina Masons discussed a fund that would collect over $14,000 in its first year,[19] and a similar fund was promoted in Indiana.[20] The educational fund established by Oregon Freemasons in 1854, which was endowed with $525 as one of the first actions of this Grand Lodge, is particularly interesting as every Mason contributed $5 to the fund, whether single or married.[21]

Though not the focus of this work, Freemasons supported public schools throughout the world. In 1813, the Masons of the Cape Colony established an educational fund to help address the "deplorable state of education" at the Cape and later established a school there (Harland-Jacobs 2007, p. 149). And in 1923, the former Consul-General of the United States at Rio de Janeiro, attributed the popularity of Freemasonry with the general public to their efforts in establish and support free public schools (Mackey, Clegg, and Haywood 1946, p. 149).

[15]This citation is drawn from Thornton (1970). In the quote by Eby provided by Thornton, Eby mentions that the role played by the Masons was possibly larger than any single religious denomination.

[16]Stillson and Hughan 1890 p. 248, Mackey, Clegg, and Haywood 1946 p. 478

[17]Moore 1846 p. 171

[18]*Ibid* p. 28.

[19]*Ibid* p. 252.

[20]*Ibid* p. 29.

[21]There was only lodge at the time in Oregon. But this level of subscription was referred to as 'remarkable' by this source. Stillson and Hughan (1890) p. 396.

## 2.4 Freemasonry in the Nineteenth Century

American Freemasonry was a social institution that played a prominent role in American life throughout the mid- to late-nineteenth and into the early twentieth century. Representing over 9 percent of adult white men at its peak in the early twentieth century, and being the basis for the fraternal orders in which nearly half of Americans participated in the late nineteenth century, the influence of Freemasonry reached to all corners of the United States.

Though Freemasonry was first established in the United States in the mid-18th century and had spread to 20 of the 26 states by 1840, the public presence of the organization contracted quite severely during the 1820s and 1830s. In particular, in 1827 the Freemasons experienced the infamous so-called "Morgan Affair" where an ex-mason, William Morgan, was supposedly killed by Freemasons to prevent him from revealing the secrets of the order. Though Freemasons were accused of kidnapping and killing Morgan, the accused were set free by a judge who was also a Freemason. This led to a national backlash against the Freemasons and membership numbers dropped precipitously. Various state legislatures tried to pass legislation banning the organization and in 1830 the Anti-Masonic Party was the only third party to run a national candidate (Tabbert 2005). The organization did not recover in terms of its membership numbers and particularly its public presence recovered until the 1850s.

In this section I describe the presence of Freemasonry in the Western United States during the second half of the 19th century in four ways. First, I estimate the importance of Freemasonry by calculating the amount of wealth invested in Freemasonry. Second, I document the spread of Freemasonry in the Western United States during this period. Third, I describe the composition of the organization. And fourth, I explore the social, political and economic roles played by this organization as suggested by the historical literature.

### 2.4.1 Measuring the Importance of Freemasonry in 19th Century America

Many authors have cited the prevalence of Freemasons among early the early presidents of the United States, as 7 of the first 20 presidents were Freemasons, as evidence of the influence of Freemasonry during the formative years of the United States. Other authors, such as Tabbert (2005), have pointed to other, more quantitative evidence of the significance of Freemasonry. Indeed, by the end of the 19th century nearly 3% of the white male population was a Freemason, a number which surged to nearly 9% at its peak in early twentieth century. Further, by the late nineteenth century, approximately 40% of American adults participated in at least one of the nearly 3,000 fraternal orders that had developed using the

Freemasonry model (Tabbert 2005).

Another, and I believe novel, way of understanding the potential influence of this organization is to look at the total amount of wealth invested by Americans in Freemasonry during this period. With initial membership fees reaching $50 or more, which was very significant given that GDP per capita during this period was around $70 nationally, joining Freemasonry required a significant financial commitment by new members.[22] Indeed, the magnitude of these fees suggests that it was an investment in the relationships and connections that they would develop through the organization (these fees are discussed in more detail in Section 2.4.3)

Using this interpretation of membership fees as an investment, I calculate the the amount of private wealth invested in American Freemasonry at different points of time in the 19th century and compare these estimates to the contemporaneous total amount of wealth in the United States.[23] As Freemasons incurred a variety of other costs in addition to this initial membership fee, these estimates should be interpreted as lower bounds of the investment in this organization.[24] These lower bounds of the share of total wealth in the United States in invested in Freemasonry are provided in Figure 2.1. Though I have estimates of the total investment in Freemasonry for every decade of the 19th century, estimates of this share are provided for only 1800, 1850, 1880, 1890 and 1900 as estimates of the total wealth in the United States are not available for other years. The percentage of the total population involved in Freemasonry is also reported for each of these years.

My estimates of total wealth invested in Freemasonry fell steadily during the 19th century while the percentage of the population in Freemasonry rose. In 1800, when Freemasonry accounted for 0.1% of total national wealth, Freemasonry was also one of the the largest private organizations. With members and lodges spread across at least 11 of the 16 states in the Union, the total amount of wealth investment in Freemasonry was equivalent to the capitalization of the largest contemporaneous corporations.[25]

[22]This estimate of per capita GDP is from Seaman (1868) and is for 1850.

[23]My estimates of the total amount of wealth in the United States are from Goldsmith (1952) which are described in more detail in the data section.

[24]In the data section below I provide a more extensive discussion of how these estimates were constructed as well as a discussion of why these should be interpreted as lower bounds.

[25]The capitalization of the largest corporation in the country, the Insurance Company of North America, was $600,000 in 1792, while my estimate of the lower bound of investment in Freemasonry in 1800 was approximately $700,000. The estimated capitalization of the Insurance Company of North America is from Davis (1917) who provides estimates of the total capitalization of the top 5 enterprises in the US in the late 18th century as follows: (1) Insurance Company of North America with $600,000 in 1792, (2) Providence Bank with $530,000 in 1791, (3) Bank of North America with $400,000 in 1782, (4) Massachusetts Bank with $250,000 in

That the share of the population in Freemasonry continued to rise while the share of the total wealth invested in the organization was stagnant or declined, indicates that the influence of this organization declined during the 19th century. Though this decline is only slight between 1800 and 1850, the drop between 1850 and 1880 and through the end of the century is quite significant. This decline is particularly apparent between 1880 and 1890 which is the period that Tabbert (2005) and others have pointed to as the period during which the influence of Freemasons began to wane.

Though the total wealth invested in Freemasonry nationwide fell to around 0.05% by the 1870s and 1880s, it was often much higher in the states of the Western United States which are the focus of this study. In Figure 2.2 I construct estimates of the share of wealth invested in Freemasonry for each county of these Western states using 1870 data and then graph the means and standard deviations of these estimates for each state. In most of these states, the average share of wealth invested in Freemasonry was at least 0.3% and was often significantly higher. Seven of the 24 states had averages exceeding 0.5% and 17 of the 24 states had at least one county with greater than 1% of the total wealth invested in Freemasonry.

### 2.4.2 The Spread of Freemasonry in the Nineteenth Century

As of 1850, Freemasonry's presence was quite limited in the Western United States. As is demonstrated in Figure 2.5, only about half of the populated area of this region had Freemasons, and the concentration of Freemasons was quite low in the areas where it was found. Indeed, there were only a handful of areas where Freemasons represented more than 1% of the population.[26] In some states, the small presence of Freemasonry can be explained by its recent arrival (see Appendix B.1 for the dates of the first arrival of Freemasonry to each of these Western States). However, in the states where Freemasonry had been established for at least a decade (i.e. Arkansas, Illinois, Indiana, Iowa and Missouri) this small presence is likely explained by the fact that 1850 was the beginning of the re-emergence of Freemasonry in America.

Between 1850 and 1860, Freemasonry experienced a relatively rapid expansion west from the states of the Midwest and throughout the West as can be seen by comparing Figures 2.5 and 2.6. However, while the expansion between 1850

1784 and (5) Hartford Bank with $100,000 in 1792. My lower bound estimate for investment in Freemasonry in New York alone is around $250,000.

[26]Note: Though data on Freemasonry in Louisiana and Texas during this period is not available, the data from 1860 suggests that the organization may have been better established in these two states than other states of the West during this period. I am making efforts to obtain these data as it will help complete the picture of Freemasonry during this earlier period.

and 1860 was primarily extensive, the expansion during the subsequent decade was both extensive and intensive. Indeed, in Figure 2.7 I show that by 1870, in addition to continuing its geographical expansion, the concentration of Freemasonry had also risen to 1% in most of the areas where the organization had been established. This decade corresponds to the period during which some historical scholars have suggested Freemasonry had its peak influence.

As is demonstrated in Figures 2.8, 2.9 and 2.10, Freemasonry continued to expand quite rapidly through the end of the nineteenth century. And by the end of the nineteenth century, there were very few parts of the Western United States where Freemasons were not present, and there were few places with less than 0.5% of the population was a Freemason.

### 2.4.3 Composition of Freemasonry During this Period

Freemasonry was in principle an inclusive association and open to all men who were monotheistic and of 'good moral character'.[27] Indeed, only men who were known drunkards, criminals or known to be abusive towards women were restricted from joining. However, though American Freemasonry included men of many faiths and many nationalities, high membership fees guaranteed that Freemasonry remained an exclusive organization into the late 19th century.

Many other sources, both historical and more recent, have discussed the restrictive nature of these membership fees. That several Grand Lodges were petitioned to lower the minimum fee for entry into Freemasonry highlights the awareness of the restrictive nature of these fees.[28] And though in some cases there were reports of members being refunded part or all of their membership fee after joining, potential members were in general discouraged from joining if the membership fees would create any hardship for him and his family (Demott 1986).

Though I am certainly not the first to mention the restrictive cost of membership, in Figure 2.3 I provide a novel way of thinking about this cost. In particular, in this figure I compare the GDP per capita in 1870 in the states in my sample to an approximate initial membership fee of $50. In practice, this membership fee varied by lodge with only the minimum fee for joining set by the state. Thus, though I provide the minimum state membership fees for a variety of states in Appendix B.2, in many cases the membership fees were above the minimum. I have selected $50 as it is an approximate average of these minimum fees, though it should be interpreted as a lower bound for reasons discussed in the data section

[27] One important exception to the 'inclusive' nature of Freemasonry were the Africa Americans who actually developed their own branch of Freemasonry during this period after facing significant racism from some fellow Freemasons. This organization is called Prince Hall Freemasonry and it is particularly vibrant today (Adam Kendall - personal communication).

[28] Various proceedings.

below.

From Figure 2.3 it is clear that Freemasonry was very expensive, though this cost varied from state to state. In nearly half of these states in 1870, the cost of joining was equivalent to the GDP per capita in that state. And in almost all states it was at least one-third of the GDP per capita in that state. And in California, one of the only two apparent exceptions, this same rule does hold as average fees were actually ∼$75. However, as the cost of joining stayed fixed in nominal terms, becoming a Master Mason become cheaper towards the end of the 1900s and has been suggested to be a contributing factor to the erosion of the exclusiveness of this organization.[29]

Despite the fact that Freemasonry was thus restricted to a relatively wealthy minority of the population, the few studies that have examined the occupational composition of its membership indicates that white collar and skilled blue collar workers dominated the Freemasons ranks and not the traditional elite. Dumenil (1984), who has the only study of the composition of a lodge in the late 19th century that I am aware of, found that white-collar jobs accounted for 75-80% with skilled workers comprising the rest in the examination of one lodge in Oakland between 1880 and 1900. Rosenzweig (1977) found a similar pattern in several Boston lodges between 1900 and 1930, where 77% of Freemasons were white collar. And though for an earlier period, Kutolowski (1982) found that the traditional elite were generally underrepresented among Freemasons in upstate New York in the 1830s. Interestingly, Parry (1998) reports that the composition of Freemasons in France was very similar to that found in the United States where membership there was dominated by white collar workers, though a few traditional elites were also members.

Instead of the traditional elite, there is an indication that the men joining the Freemasons were more likely to be those that were upward mobile and had political aspirations. Though joining Freemasonry for mercenary purposes was discouraged in official rhetoric, Rosenzweig (1977) finds evidence that men joined Freemasonry for economic self-interest and Weber (1946) similarly observed that 'business opportunities were often decisively influenced' by membership in Freemasonry. And with regards to political activity, in addition to the variety of examples of prominent politicians who were Freemasons, Kutolowski (1982) provides evidence that Freemasons in upstate New York in the 19th century were indeed very politically active. Indeed she finds that 55% of all county political leaders were Freemasons during the period of her study (1821-1826).

[29] This cost of joining has stayed fixed in nominal terms until the present. Thus, while there are a variety of new types of fees that have been introduced, it is significantly cheaper today. Some Masons who were involved in the administrative side of the organization told me that this is a significant issue as they are often forced to fund their organization from existing endowments as membership fees are too small.

While there is relative agreement in the historical literature on the occupational composition of Freemasonry during this period, there is a bit more disagreement with regards to both the representation of different religions and different immigrant groups within the ranks of the Freemasons. Indeed, in her study of Freemasons in Oakland, California, Dumenil (1984) found that the immigrants were significantly underrepresented and Rosenzweig (1977) found that the Bostonian Freemasons in the early 20th century were socially homogenous with little or no Catholics represented.

However, in his detailed examination of Freemasonry in 19th century San Francisco, Fels (2002) finds that native- and foreign-born protestants as well as Jews were well represented. With regards to immigrant involvement in Freemasonry, there were three lodges that each functioned entirely in a different language (one each of French, Germany and Italian). And though Catholics may have been underrepresented, he finds that Jews were particularly well represented with 12% of San Francisco Freemasons being Jewish despite despite only accounting for 7-8% of the San Francisco population. And in general, these lodges were not that segregated with several prominent Freemasons being particular vocal about encouraging integration.[30] Fels indicates that a similar participation rate of Jews was true throughout the United States.[31]

### 2.4.4 Freemasonry as an Economic, Political and Social Institution

> "In the late nineteenth century... [Freemasonry] offered sociability, relief in times of distress, as well as possible financial and political advantages." - Dumenil (1984) (p. xii)

Men joined Freemasonry for a variety of different reasons. While they joined for the tradition, the pageantry and to socialize, they also joined because of the perceived private benefits and opportunities that they would gain from being part this sophisticated local, state, national and international exclusive organization. In this sub-section I review the extant literature that explores how Freemasonry may have providing economic and political benefits to its members as well as the role that it played in the development of civil society in the Western United States.

One common perception is that people joined the Freemasons because it would give them an advantage in finding employment opportunities. And though joining for what might be perceived as mercenary reasons was frowned upon, Masons did have a reputation for "sticking together". Dumenil (1984) reports that there was a common understanding that Masons would hire another Mason

[30]Jews were, however, excluded from a few peripheral orders (Fels 2002).
[31]See Fels (2002), footnote 25.

if they had to choose between two candidates as "employing a fellow Mason not only helped a brother, but also was supposed to assure the employer of an honest, upright employee" (p. 22). Similarly, as there were many bankers that were Masons, it is likely that the Freemason network would have benefited men starting new business enterprises. Indeed, Braggion (2009) provides compelling evidence that the Freemason network did help alleviate agency problems between lenders and borrowers among small firms in London.

Another important private benefit was the "risk sharing" that was provided by Freemasonry. Indeed, one celebrated aspect of Freemasonry was that a man's family would be provided for if he were to die.[32] This form of insurance was likely quite important in the frontier states of the American West during this period. And the importance of this aspect of Freemasonry to the Masons themselves is demonstrated by the fact that this was one of the two areas for which fund raising was often done (the other one being education as discussed above).[33] That this insurance was important during this period, and perhaps an important attraction of Freemasonry, is that this type of insurance was the *raison d'être* of many of the fraternal orders that emerged from the Freemason model in the late 19th century.

Though these direct economic benefits from joining Freemasonry certainly played a role in some parts of the American West during the 19th century, it is likely that the social aspects of Freemasonry during this period were *at least* as important. In many communities, the only places for men to congregate were either the saloon, the brothel, the church or the Freemason lodge. Thus, for those that did become Freemasons, the lodge was often the only place where it was possible to form meaningful relationships. Indeed, Haywood (1963) argues that Freemasonry "became all things to its members, a lodge, a religious center, a social rendezvous, a league for self-protection, and a sieve with which to sift the wheat from the chaff among the American adventurers, miners, trappers and traders."[34] And friendships were often solidified after the weekly meetings as as Freemasons often gathered after their meetings for food, drink and conversation (Dumenil 1984).[35]

Thus, the bonds that were formed between Freemasons were often stronger than the trust with "members of their own families, the people with whom they prayed, their political allies, and even their business partners" (Burt 2003).[36]

[32]The cost of his funeral was often covered by the Freemasons as well.

[33]Various Proceedings.

[34]This is originally from Haywood (1963) p. 206. However, my source for this quote is De Los Reyes and Lara (1999) as the original was not available to me.

[35]Interestingly, it seems that the lodge likely actually served as a complete substitute for the saloon in this way.

[36]Burt focuses on Freemasonry among mining towns in Britain. However, his conclusion is very similar to that reached by many historians of American Freemasonry and I choose to quote his words as I feel that they are particularly apt.

While these bonds were partly a by-product of participating in the same secret rituals, there was a much more pragmatic aspect to this trust. Indeed, a Freemason would me very unlikely to betray a fellow Freemason as expulsion from the order would entail the loss of both the money he had invested in the organization as well as the many advantages that membership offered. Thus, the Freemason network may have functioned to reduce the principal agent problem in the same way as was found by Greif (1993) among early trade networks in the Mediterranean, an idea which has also been suggested by both Burt (2003) and Braggion (2009) for Freemasons.

In addition to possibly reducing principal agent problems, Freemasonry also played an important role as tool for the dissemination and sharing of information, which is essential for economic development, in at least some communities. Clark (1936), who examined Freemasonry in communities in the American West in the 19th century, highlighted the importance of Freemasons' role in this:

> "We agreed to meeting every Saturday night... we decided that any ideas concerning the country we were in, which might come to us, news of any mines we might discover, or any information which might be beneficial to the brethren, Masonically or financially, would, at the next meeting, be given to the Masons, there assembled." (Clark 1936, p. 56-57)

In England, information about the profitability potential of all mines throughout the country spread first through Masonic networks (Burt 2003). It is likely that these networks played a similar role in agricultural areas, though to my knowledge the historical Freemason literature has not explored this issue.

And these networks were not only local. As a Freemason was welcome in any lodge anywhere in the country, and indeed throughout the world, these networks were both national and international. A migrant Freemason would typically find support, both financial and social, within a local lodge.[37] Burt (2003) goes so far as to argue that men may have joined the Freemasons to take advantage of its international networking links and considered Freemason membership to be an 'emigration asset'.

One externality of these bonds and networks, that has been suggested by some authors, is that Freemasonry may also have played an important role in the development of civil society and organizational life throughout the country during the country. De Los Reyes and Lara (1999) demonstrate the role that Freemasonry played in the development of civil society in several communities and suggest that a similar effect was found throughout the United States in the 19th century. The 3,000 fraternal orders, which represented over 40% of the adult population, which

[37] Indeed, there are apocryphal tales of men posing as Freemasons - or perhaps Freemasons who were expelled from the order - traveling from community to community to take advantage of exactly this benefit of the organization.

emerged in the late 19th century based on the model developed by Freemasonry indicate that this was likely true (Tabbert 2005).

Another possible externality is that Freemasonry, which was a rational organization at a very early point in history, facilitated the development of the modern corporation in the United States. As mentioned above, the amount of wealth invested in Freemasonry at the end of the 18th century made it one of the largest enterprises in the county. And by 1850, almost all the Freemason Grand Lodges in the county had established a very similar form for recording financial and membership information. Thus, in addition to possibly facilitating enabling relationships between ambitious and driven, the Freemason organization gave men the impetus and opportunity to explore new hierarchical and organizational technologies on a national scale. This suggests that Freemason may have played a role similar to that attributed to the railroads by Chandler (1977).[38]

## 2.5 Data

### 2.5.1 Freemasonry Membership Data

American Freemasonry is a fully federated system with the lodges of each state reporting solely to the Grand Lodge of that state. Each year, each of the Grand Lodges would hold a state conference and representatives from each lodge were expected to attend this conference.[39] After these conferences, the Grand Lodges would publish a volume, typically called the Grand Lodge Proceedings of [name of state], which would report the discussions that were held during the conference and correspondence with other Grand Lodges throughout the United States and the world.

Beginning in 1850, these Grand Lodge Proceedings began to publish either lists of the members of each lodge in the corresponding state or memberhsip statistics for each of these lodges.[40] While Grand Lodges could not form in a state until that state had officially entered the Union, lodges would still form in those

[38] Burt (2003) similarly suggested that the "skills and personal attributes needed to run a harmonious lodge were similar to those required in a successful business enterprise" and that the Freemason lodge demonstrated how a variety of different professions could coalesce in the pursuit of commonly agreed on purposes. He also suggests that by the mid-19th century that the Freemason networks had established the institutional foundations for the emergence of national and international "communities" and that they had an "unparalleled capacity" to bridge very geographically dispersed communities.

[39] The cost of transportation, lodging, etc. for these conferences would typically be subsidized by the home lodge.

[40] A few states have this membership data available before 1850: Rhode Island has data going back as far as 1790, Tennessee has data starting in 1820, Virginia starts in 1800, etc. However, all of the state have some sort of membership data available by 1850.

areas under the jurisdiction of one of the other Grand Lodges.[41] Thus, as census data is available for some areas before they become states (e.g. Dakota Territory, Indian Territory, Nevada Territory, etc.), Freemason membership data can also be constructed for many of these states from the Grand Lodge Proceedings of other states.

Using the historical Proceedings from each of the Grand Lodges, I was thus able to collect lodge level membership data for each state from 1850 to 1900. Though data is available for all states, I focus on the Western United States in this study. In particular I include all states west of the Mississippi plus Illinois, Indiana and Wisconsin.[42] The Western United States provides a very interesting environment for studying the role of a social institution such as the Freemasons on development, as much of this region was just beginning to develop during this period. Further, this sample includes a variety of states at different stages of development. In principle I would eventually like to include all the states in my study, but time limitations have made this impossible for the current study.[43]

I collected this membership data decennially and county-level estimates of the number of Freemasons were calculated by aggregating all Lodges within a county, as this allows me to match this membership data to the most disaggregated socioeconomic data available during this period from the US census (see below for more details on this data).[44] Typically, most of the Proceedings have a statistical abstract or appendix that lists at least the name of the lodge, the number of the lodge and the number of Master Masons. In many cases these appendices also contain the geographic location of the lodge, either city or county, in that year. However, this geographic data is not available for all years in every state.

For this present study, lodge location was usually determined using the information published in the 1900 Proceedings. However, for several states, the location of lodges that became extinct before 1900 are not reported.[45] For these

---

[41]This is the mechanism which allowed for Freemasonry to spread as individuals were not allowed to spontaneously establish lodges. As an example, the early lodges in the state of California were under the jurisdiction of the states of Illinois and Missouri.

[42]I include these latter states for two reasons. First, is because these are the location of the second wave of industrialization in the United States, which began during the period in question, and I think it is important to understand the role that Freemasonry may have played in this. Second, is because this is the area where I grew up - I actually had one of the earliest Masonic orphan homes in my hometown.

[43]In theory I could just send these membership data to a company for digitization. In practice, idiosyncrasies in each publication of the Proceedings for each state makes this effectively impossible.

[44]In principle, I actually have city/village level Freemason membership numbers available. However, these data are not useful for the present study as there is no easily accessible, geographically comparable source of socioeconomic data for these cities.

[45]Several states continue to publish the location of extinct lodges in each revision of the Proceedings.

extinct lodges, the location is determined in two ways: First, when available, the location is extracted from one of the previous Proceedings. Second, for lodges where this is not possible the location of the lodge is sometimes imputed by matching the name of the lodge to cities in that state (lodges were often named after the town where it was located or vice versa). This is not always possible as the lodges often either take on Masonic names that do not match up to city names or the same city name appears in more than one location in a state.

### 2.5.2 Share of Wealth Invested in Freemasonry

In this sub-section I describe my calculations of the share of wealth invested in Freemasonry. The estimates of the investment in Freemasonry as well as total needed for these calculations are described separately here.

#### Investment in Freemasonry

Lower bound estimates of the share of total wealth invested in Freemasonry are reported in Figures 2.1 and 2.2. Figure 2.1 reports my estimate of the national aggregate and Figure 2.2 reports the means and standard deviates of county data for all states in my sample separately. Here I discuss my approach for calculating this lower bound estimate and discuss why it is appropriate to treat this as a lower bound.

Though there are a variety of approaches that I could have used to estimate this quantity, I chose to calculate a lower bound estimate by comparing the total wealth in a geographical area to an estimate of the total amount of money spent on initial membership fees alone. Thus, my lower bound estimate of the total investment, $I^F_{LB}$, in Freemasonry is simply

$$I^F_{LB} = (Membership\ Fee) \times (\#\ Masons)$$

and then I divide this quantity by the total wealth (discussed in the section 'Total Wealth' below). As data on the membership fees charged by individual lodges is unavailable, I used a constant value of \$50 across all states and periods in calculating this lower bound as discussed immediately below.

In practice, membership fees varied significantly by state, and by lodge as the Grand Lodge of each state only has the authority to set the minimum membership fee permitted. Thus, though I was able to collect some data on minimum membership fees, which I provide in Appendix B.2, the actual membership fees charged by individual lodges was often significantly different. Further, the data that I do have on membership fees are from the late 19th century, which are anecdotally *less* than fees charged earlier in the century. The value of \$50 that I have used was selected to be a rough average of the lodge membership fees during at least the first 70 years of the 19th century, though fees likely varied from $30-200.

However, this estimate should be treated as a lower bound of the total investment in Freemasonry for several reasons. First, many Masons joined one of the sister organizations of Freemasonry, which all had additional initiation fees and annual fees. Two prominent examples are the Knights Templar and the Royal Arch Masons. The the Knights Templar, in which between 10-30% of Freemasons participated, cost an additional $75 in New York, $100 in Chicago and $230 in San Francisco (Dumenil 1984, Hope 2005). And though the fee and participation rate likely varied by state, Royal Arch Masonry in California had initiation fees of $30 and accounted for nearly one-third of Freemasons in California.(Royal Arch Masons of California 1882).[46]

Second, the variety of vestments that were required for Freemasonry and its various suborders were often very expensive. Though I was unable to locate any reliable data on the cost of clothing for the main organization, data on the cost of vestments for the Knights Templar is available give an idea of how significant these costs might be. Before a mason could become a Knights Templar, a uniform costing around $40-100, exclusive of the cost of a horse, was required in addition to the initiation fee.[47] Indeed, these uniforms were so costly that the total value of uniforms worn in a single Templar parade was $1,000,000, which is roughly equivalent to $100,000,000 in current dollars.[48]

Third, this figure does not include the number of women who were members of the Eastern Star nor estimates of the investments that they may have made in that organization. There were nearly 50,000 women in this organization in 1890, which was one-ninth of the total men in Freemasonry during this time (Stillson and Hughan 1890).

In addition to these fixed costs, which I argue can be interpreted as an investment, Freemasonry had a variety of non-trivial annual and more regular fees. Annual fees paid to both the local and the state Grand Lodge were likely at least $5 in most states, which was not an insignificant cost as the GDP per capita in many Western states was below $100 (Figure 2.3). And a variety of Freemason social events, such as balls and the like, would also require further contributions from the Masons.

Further, charitable giving was typically a very important responsibility as evidenced by their frequent discussion in various Grand Lodge Proceedings. The beneficiaries of these charities varied but the most popular causes were for the support of widows or orphans (through orphanages), educational funds or the

[46] 3,827 of the 12,132 Freemasons in 1880 were also Royal Arch Masons.

[47] Hope (2005) reports that this uniform cost around $100 in California. An article in the Washington Post provides a slightly lower estimate of $40 (*The Wasington Post*, September 27, 1889, pg. 1). Dumenil (1984) reports a cost of $71.

[48] *The Wasington Post*, September 27, 1889, pg. 1.

support of sister lodges in the state or elsewhere in the county in a time of crisis. These donations were typically a few dollars per Mason per year, which was often a significant amount of money during this time. Indeed, one Templar picnic in New York raised $10,000 (approximately $1 million in present dollars) for one of their charitable funds.[49]

### Estimates of Total Wealth

Here I discuss the estimates of total and county-level wealth that were used for calculating the share of wealth invested in Freemasonry. Estimates of the total wealth in the United States used for Figure 2.1 are drawn from Goldsmith (1952). However, the estimates of county level wealth used in Figure 2.2 are my own and require some justification.

Though the total amount of wealth is not directly available in the historical U.S. censuses, Goldsmith (1952) provides estimates of the total wealth in the United States for several decades. His work, which draws on the work of Kuznets (1946) and others uses data available in the census to generate estimates of the total wealth. Though these estimates require several assumptions, and in particular a relatively strong assumptions about the relationship between value of farm real estate and agricultural capital, they are to my knowledge the best available.[50] As Goldsmith (1952) only provides estimates of the total wealth in the United States for 1800, 1850, 1880, 1890 and 1900, estimates of the share of wealth invested in Freemasonry is restricted to these years (Figure 2.1).

Estimates of total county-level wealth are not available for this period. However, there are two estimates of capital available at the county-level in each wave of the census. The first is an estimate of the physical capital in manufacturing: 'Total capital invested (in dollars) in manufacturing' which includes capital invested in both buildings and machinery.[51] The second, 'Present cash value of farming implements & machinery', is an estimate of the physical capital in agriculture. I use these two measures of capital to estimate the total wealth in each county.

My approach to estimating county-level wealth is necessarily rough as these two variables only cover a fraction of the total wealth in a county. In particular, the national aggregate of these two variables represents approximately one-sixth of the total wealth throughout the second half of the 19th century.[52] In lieu of

[49] *The New York Times*, October 5, 1870, pg. 2).

[50] Kuznets (1946) and Goldsmith (1952) generate an estimate of the total amount of agricultural capital in land by multiplying the total value of farm real estate by one-fifth, which was the average of improvements to real estate in 1900.

[51] That it includes total capital in manufacturing is indicated by Kuznets (1946) who calculates the total value of manufacturing buildings as a ratio of this amount.

[52] Goldsmith (1952) reports total wealth in the United States in 1900 at $55 billion while the aggregate of these two variables in the census data is $10.6 billion and the corresponding values

a more precise way of estimating county-level wealth, I use this "rule" of a ratio of 6:1 between total wealth and the amount of capital represented by these two variables to derive rough estimates of total county-level wealth.

Thus, my estimate of the total wealth in a county is simply the sum of the agricultural and manufacturing capital reported in the U.S. Census for that county scaled up by a factor of 6.[53] However, as the true relationship between total wealth and these measures of capital is likely to vary by county, this is necessarily only a "back of the envelope" estimate.

Though this approach for estimating county-level wealth is thus quite rough, I believe it is adequate as I use it only to provide an approximation of the magnitude of the investment in Freemasonry. And though there are other possible approaches to estimating total wealth at the county-level, these approaches will in general require very strong assumptions. In particular, assessing the value of agricultural land is particularly difficult, though essential for deriving precise estimates of county level wealth.[54]

### 2.5.3 Dependent Variables

The educational variables that are used in this analysis are drawn from two key sources. The first, which are used primarily for the cross-sectional analysis, are county-level aggregates available from the historical U.S. censuses. The second are the 1% IPUMS historical samples which are used for the panel estimates.

#### Census County Aggregates

The cross-sectional analysis focuses on three types of dependent variables that are available in the historical U.S. censuses.[55] The first type is school enrollment data which is available for 1850, 1870, 1890 and then frequently throughout the 20th century. Second, in 1870 there is data on the level of taxation at the local (town/city), county and state level. Finally, which is only available for 1850, is a variety of detailed data concerning the structure of the funding for private versus public school as well as the total number of teachers.

Though educational enrollment data is available in many of the censuses, it differs slightly between each of the censuses. Here I discuss the type of enrollment data that is available in 1850, 1870, 1890 and 1940. Though in some

for 1880 are $22.4 billion and $3.2 billion.

[53] I use the extracts from the historical U.S. Censuses developed by Michael Haines. *ICPSR 2896.*

[54] Kris Mitchener, personal communication.

[55] I use the extracts from the historical U.S. Censuses developed by Michael Haines. *ICPSR 2896.*

years it is possible to calculate enrollment separately by gender or by race, neither the Freemasonry historical record nor the simple theoretical framework have any separate predictions about these quantities. Thus, while I mention them here as they be of interest for future analysis, the focus of my quantitative analysis of enrollment will focus on gross enrollment.

In 1850, the data on school enrollment is available in two forms. First it is possible to separately identify the number of students in private and public schools using the following two variables: 'Number of pupils at academies and other schools' and 'Number of pupils in public schools'. While this would be an interesting distinction to study, only 13 counties in in my sample have any data on these academies, and it is thus impossible to study them using data on only the Western United States. Second, it is possible to calculate the total enrollment rate among whites, which will be the variable I will focus on below - though it is possible to separately identify enrollment rates among (white) girls and boys using these data.

For 1870, it is only possible to calculate the gross enrollment rate. I calculate this gross enrollment rate by dividing the 'Total persons in school' by the total number of youth ages 5-18. Though there is relatively extensive data on attendance among several different groups (e.g. 'native-born', 'foreign-born', 'colored'), the denominator - i.e. the total number of youth of these different groups - is not available.

For 1890, it is possible to calculate school enrollment in three different ways. First, it is possible to calculate the total enrollment rate in common schools among whites by dividing the number of white men and women in common schools by the total native and foreign whites youth.[56] Second it is possible to separately calculate enrollment among white men and white women. Finally, it is also possible to calculate the total enrollment among 'colored' youth during this period.

For 1940, it is possible to separately measure the enrollment rates among youth of elementary school, lower secondary, upper secondary and college age separately. Similar data is available from 1910-1930 though I do not include it in this analysis as the results are generally quite similar - and in the few cases where there are differences I mention them.

In 1870, there is data on the total amount of state, county and local taxes paid by the residents of each county. While it is unclear how local taxes are being used, and it is likely to differ across counties, the public good aspect of Freemasonry predicts that the ratio of local to state should be higher in counties with more Freemasons.

Though the number of states with both economic and Freemason data is rather limited for 1850, the data on education is the most extensive in 1850 along

[56]The data do not allow me to separately identify the number of foreign versus native whites enrolled as in this case I lack the appropriate numerator.

two important dimensions. First of all, the amount of income for public schools, private schools and colleges from a variety of sources - including taxation and public funds - is available. Second, in 1850 there is data on the total number of teachers in public schools, private schools and colleges so that it is possible to calculate the teacher to pupil as well as the teacher to total youth population ratio.

#### IPUMS 1% Sample

While the county-level aggregates provide data on the universe of children in calculating enrollment rates, the definition of child enrollment is not constant throughout the 19th century and is thus not appropriate for panel data estimation approaches. However, IPUMS-USA has made available 1% samples from each of the 19th century censuses that allow for the creation of a child enrollment panel (Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek 2010).

These data are used to construct a panel of county-level youth enrollment among children ages 5-18 decennially for 1850-1900. Though these data are more sensitive to measurement error as it is only a 1% sample, the panel structure allowed by this data provides an important tool for verifying the robustness of my cross-sectional results.

### 2.5.4 Demographic and Economic Data

Throughout my analysis I include a variety of demographic and economic variables that are drawn from the historical US censuses. The meaning of many of these variables is quite clear. Indeed, the total size of the population, the number of men, the number of people living in cities of different sizes, etc. are standard census quantities. However, throughout this analysis I include four separate measures of the income and wealth of these counties. I feel that a short discussion for including these quantities separately is warranted.

Two separate measures of wealth are available throughout the census data of the late 19th century. The first of these is the 'amount of capital invested in manufacturing establishments', which I call $K^M$. The second is the 'value of farming implements & machinery', which I call $K^A$. It would be inappropriate to aggregate these measures together for the estimation as each of these variables only measure a fraction of the total capital in a county and this fraction is unlikely to be the same across the two sectors. However, as long as the ratio between the measured amount of capital and the true amount of capital in each sector is constant across a state, and I include both in my regression, then state fixed effects will capture this ratio.

The data on income is slightly more limited than that of wealth.[57] In particular, while data on manufacturing output - the 'value of annual product in manufacturing establishments' - is available beginning in 1850, there is no aggregated data on agricultural output available until 1870 ('total (estimated) value of all farm productions, including betterments & addition to stock'). In practice, it is possible to extend the approach of Tucker (1843), Seaman (1868) and Easterlin (1960) to county level data and simply multiply the agricultural production of each county by the corresponding prices (available in the two first references). However, after having done this for one state I decided that the effort involved - as all the agricultural production data needs to be hand entered due to the low quality of the available scanned agricultural censuses - was significantly less than marginal benefit of having a measure of agricultural output for these earlier years.

A possible concern is that the fraction of the real wealth or income that is measured by these variables may vary with the some of the either right-hand side variables, such as the number of Freemasons or the total population. This type of measurement error would lead to biased point estimates. Unfortunately, I have been unable to identify any sources that discuss in detail the collection of these variables or possible concerns with using them.[58]

### 2.5.5 Shifting County Borders

County definitions in many states changed significantly during the second half of the twentieth century. Thus, as my lodge location data is primarily from 1900, I needed to account for the possibility that a lodge was actually in a different county in an earlier year. In particular, I needed to find the relationship between counties in 1900 and each of the other years for which I have Freemason data: 1850, 1860, 1870, 1880 and 1890.

My approach for addressing this problem proceeded in two steps. First, historical GIS maps of county borders in the United States were obtained from the National Historical Geographical Information System.[59] I then used the GIS ArcMap software to find counties were border definitions likely occurred by comparing the 1900 county map with the county maps for each of the years. In each case, I calculated the intersection of two county maps, using a one-mile tolerance, which would identify any changes in county definitions that were more than one

[57] Estimates of both forest production are available beginning in 1870 and estimates of mining production are available beginning in 1880. However, this data would also need to be hand entered as it is not available in any digital form that I know of. This omission is unlikely to have a large negative impact on my results.

[58] I did on several occasions contact the US Census Office. While they were very helpful, there did not seem to be very much institutional knowledge about the specific of the collection of these variables. And returning to the old census reports did not reveal very much information.

[59] http://www.nhgis.org/

mile.

In the second step, in order to correct for mistakes in this matching process, I checked all the likely matches found in ArcMap by hand. In particular there are many cases where county borders between the two county maps would shift by three or four miles. Or in some cases there were changes where a frontier county would 'lose' much of its frontier territory - though it is unlikely that this there was any significant population in this frontier territory in earlier years. This second step was thus subjective and relied a bit on my 'best judgement'. But it was essential, as there were many 'mistakes' in the first matching step. Also, it was done entirely before any economic analysis was done so it is unlikely to affect the results.

### 2.5.6 Potential IVs

While there are a variety of variables available in the census that one might think would be appropriate, in all cases these variables have either an implausible first stage or exclusion restriction. As examples I will show how three leading possibilities are not appropriate: (1) Share Catholic [*strong first stage, implausible exclusion restriction*]: strong and significant first stage as the Freemasons and Catholicism had an antagonistic relationship during this period, but the Catholic Church also had an independent impact on public education as they were actively promoting and trying to secure funding for private, religious schools. (2) Share Foreign [*no first stage*]: there is no evidence that foreigners received preferential treatment or were excluded from the Freemasons. (3) Blue Laws [*no first stage*]: While some lodges were completely dry and encouraged alcohol abstinence among their members, members of other lodges would gather together after meetings to drink together - and this type of variation is found *within* states as well as across. Similar argument can be applied for every other variable in the US census as well as other variables that would require great expense to collect.

## 2.6 The Impact of Freemasonry on Educational Enrollment

Here I use the county-level data on educational enrollment available in the US census to estimate the impact of Freemasonry on education. The analysis is divided into two major sub-sections. The first uses cross-sectional data to examine the impact of Freemasonry on education during two historical periods: (1) the late 19th century and (2) the early- to mid-20th century. Then, in a second subsection, I use a panel of education and Freemasonry data to verify the robustness of these results to the possibility of unobserved heterogeneity, endogeneity and autocorrelation.

Throughout this analysis, I will focus on estimation of a very simple relationship between educational enrollment and a variety of socioeconomic variables. In particular, I will model the educational enrollment in a county as

$$y_i = X_i'\beta + \epsilon_i \tag{2.1}$$

where $i$ indexes counties, $y_i$ is the educational enrollment in county $i$, $X_i$ contains a variety of socioeconomic variables and $\epsilon_i$ is the error process. This simple parameterization is based on the model used by Goldin and Katz (1999) in their examination of county-level data from Iowa in 1915 though I augment the set of socioeconomic variables to include my measure of Freemasonry, the number of Freemasons per capita. Also, my dependent variable differs slightly. While they focus on high school and college enrollment, which is only available in later years, I focus on total enrollment among youth ages 5-18. I also include several additional measures of wealth that are only available in early census data as well as the percentage of the population that is male and over age 21. As only adult (21+) males are eligible to becomes Masons, including this latter control ensures that my measure of Freemasons is not just functioning as a proxy for adult males who are more likely to be economically active than other segments of the population.

All analysis done here is at the county-level and all regressions include state fixed effects.[60] While it is plausible that Freemasonry also had an effect on state level education policies as suggested by the historical literature on Texas, most of the funding for education was done at the local level as discussed above. Further, it would be very difficult to measure the influence of Freemasons in explaining state-level differences as Freemasons were ubiquitous in state, as well as national, politics.

### 2.6.1 Cross-Sectional Estimates

In this first sub-section, I use cross-sectional data to explore the the potentially changing relationship between Freemasonry and educational enrollment during the early years of the development of the American West. I begin this analysis by examining the relationship between Freemasonry and total youth educational enrollment in 1870 and 1890. By 1870, Freemasonry's presence was already well established in the Western United States and the organization continued to expand rapidly through 1890. Here, I find a strong positive relationship between Freemasonry and education. And this impact seems to be similar in both agricultural and manufacturing communities.

[60]In the case of the fixed effects panel data estimates, there are no state fixed effects as it is not possible to identify the effect of variables that are time constant.

In order to test for the possible lasting impact of this social institution, I then look at the relationship between nineteenth century Freemasonry and enrollment rates in the early and mid-twentieth century. As it is possible to separately identify primary, secondary and tertiary educational enrollment in the twentieth century census data, I am able to examine both the potential lasting impact of Freemasonry on primary educational systems as well as the role that Freemasonry played in the second wave of educational expansion in the United States. While there does not seem to be a significant lasting impact on primary education, which is likely explained by the near universal primary education at this point, I do find evidence of an impact on secondary and tertiary education. This effect seems to be most pronounced among potential upper secondary students.

### Education in the Late 19th Century

Between 1850 and 1870, Freemasonry spread very rapidly in the West and into the Midwest, and by 1890 the organization was present in almost every county of the Western United States. Here I examine the contemporaneous relationship between Freemasonry and education during this period which I believe to be the height of Freemasonry influence. As both measures of social homogeneity used by Goldin and Katz are available for this period, I will be able to directly compare the 'homogeneity effect' observed by these authors to the potential Freemasonry effect.

Table 2.1 presents the results from my analysis using the education, socioeconomic and Freemason data from 1870 and 1890. For 1870 the dependent variable is the attendance rate of youth ages 5-18; and in 1890 the dependent variable is the attendance rate among all white children ages 5-20. In Columns (1) and (6), I reproduce the analysis of Goldin and Katz (1999) and find somewhat similar results as these authors using my data. In particular, in the 1870 data I find that measures of wealth as well as one of Goldin and Katz' proxies for social capital - social homogeneity as measured by the proportion of the population of native heritage - have a significant positive effect on educational outcomes. Interestingly, in the 1890 data, this same measure of social capital has a strong positive impact while their other proxy for social capital, the share of the population in small towns, has a large negative impact. This unusual result for their second social capital proxy could be driven by the fact that my measure of town "smallness" is much larger than theirs - however, the reversal of the sign with a significant point estimate is a peculiar result.[61]

In columns (2) and (7) I augment these regressions to include the per capita

[61] Twentieth century data allows a much finer disaggregation of population. And in particular, these authors observe the share of the population in towns smaller than 1,700 persons, 1,700-2,500 persons, etc.

concentration of Freemasons. In both the 1870 and 1890 data I find that the number of Freemasons per capita has a large and very significant impact on enrollment outcomes. Interestingly it seems that the introduction of Freemasons per capita has only a weak effect on the point estimates for Goldin and Katz' proxies for social capital, suggesting that Freemasons are not simply a measure of social homogeneity.

In the data from 1890, the impact of Freemasonry on education is of roughly the same magnitude as that of social homogeneity. Indeed, while the estimated coefficient on Freemasonry is about thirty times larger than that for social homogeneity, the standard deviation of the social homogeneity variable is about thirty times the size of that for Freemasons per capita. However, in 1870 the total impact of Freemasonry is about twice that of social homogeneity as the coefficent is still thirty times larger while the standard deviation is only one-fifteenth that of the measure of social homogeneity.

In columns (3)-(5) and (8)-(10) I disaggregate the counties by the type of economic activity, in order to explore where the impact of Freemasonry was the strongest, if anywhere in particular. To do this, I create three groups of activity: The first group, which I have designed to be the 'mostly agricultural' counties are counties where the share of output from manufacturing is at most one-third, are in columns (3) and (8). The second group is the 'mixed' group and is counties where the share from manufacturing is between one-third and two-thirds. The final group is the 'mostly manufacturing' counties where the manufacturing share is at least two-thirds.

As areas that are primarily agricultural have the the least dense populations, and are thus more likely to have difficulty in providing common goods, it is unsurprising that Freemasonry is positively related to educational attainment in these areas in both 1870 and 1890 as is demonstrated in columns (3) and (8). Interestingly, in these areas in 1870, it seems that Freemasonry is the only measure of social capital that has a positive impact on educational outcomes.

While Freemasonry also has a positive relationship in the two groups of more manufacturing intense counties, the effect is not universally significant. Indeed, in 1870 we see a significant effect among only the most manufacturing intense (column (5)), and in 1890 we only see a significant effect among the 'mixed' counties (column (9)).

### Lasting Impacts on Education

In order to explore the lasting impacts of Freemasonry on the education system, I examine the relationship between the concentration of Freemasons within counties in the 19th century and education in the twentieth century. In particular, I hope to explore the possible role that the social institutions established by Freemasonry in the 19th century had on the "second transformation" of American education, which occurred from 1910-1940 and saw the massive expansion of the American secondary education system (Goldin and Katz 2008). The results I present here are for educational outcomes at the end of this second transformation in 1940, though the results are similar for 1910-1930. In the few cases where there are significant differences between these years, I will mention it below. For this analysis I will use the concentration of Freemasonry in 1870, as this is often celebrated as the apex of the political influence of Freemasonry, though the results are similar using other 19th century Freemasonry data.

In Table 2.2 I again use a parameterization similar to Goldin and Katz (1999) to explore the relationship between the concentration of Freemasons in 1870 and education outcomes in 1940. This table is divided into four sections, where each section considers the impact of Freemasonry and other factors on the educational outcomes among the four following age groups: primary, lower secondary, upper secondary and tertiary.

In the first column of each of these sections, I use a parameterization that considers the inclusion of only 1940 socioeconomic variables. Interestingly, Goldin and Katz' proxy for social capital, the proportion of population that is native born, does not seem to have a significant effect on the educational outcomes among any of the four age groups considered. This effect is only found in the 1940 and not in any of the previous three decades of this "second transformation" where social homogeneity is still found to have a significantly positive effect.

I then introduce socioeconomic variables from 1870 in the second column of each section, and finally my measure of the concentration of Freemasonry in 1870 in the third column of each section. While Freemasonry has only a weakly positive effect on primary education, there is a strong positive impact on all other levels of education. Here, however, the effect of Freemasonry is slightly weaker than social homogeneity as the ratio of the standard deviations of social homogeneity and Freemasonry is about fifteen-to-one.

### 2.6.2 Panel Data Estimates

The above analysis focused on the estimation of equation (2.1) using cross-sectional data. However, consistent estimation of equation (2.1) requires that $X_i$ and $\epsilon_i$ are uncorrelated, i.e. $Cov(X_i, \epsilon_i) = 0$. However, this is unlikely to be the case as there are unobservable characteristics of these counties that are likely to affect education, Freemasonry membership and some, if not all, of the other socioeconomic variables contained in $X_i$. In particular, it is likely that

$$y_i = X_i'\beta + c_i + \epsilon_i \tag{2.2}$$

is a more accurate representation of the process determining $y_i$, where $c_i$ represents unobservable county-specific characteristics, and that $Cov(X_i, c_i) \neq 0$. In this case, the cross-sectional estimates presented in the previous section are inconsistent. Indeed, consistent estimation of $\beta$ using cross-sectional data would require the availability of either a suitable proxy or a plausible instrument for $c_i$, neither of which are clearly available (Wooldridge 2002).

However, using the 1% IPUMS data available for each decade from 1850 to 1900 to estimate educational enrollment, I construct a county-level panel data set which allows consistent estimation of (2.2).[62] In particular, this panel data allows me to estimate

$$y_{it} = X_{it}'\beta + c_i + \epsilon_{it} \tag{2.3}$$

where $i$ indexes counties, $t$ indexes decades and $c_i$ is the unobservable county effect. In this simple panel data model, consistent estimates of $\beta$ are available directly if $X_{it}$ are strictly exogenous conditional on $c_i$, or $\mathbb{E}[\epsilon_{it}|X_{it}, c_i] = 0$.

I estimate (2.3) using three different panel data sets of educational enrollment. The first of these data sets has the most coverage in terms of time, spanning from 1850-1900, but contains the least geographic coverage as it uses the earliest available county definitions which tend to be more aggregated than later counties (T = 6, N = 451). The third data set covers the fewest years, only 1870-1900, but offers the greatest geographical variation (T = 4, N = 1034) while the second, which covers 1860-1900, is a balance between geographical variation and years covered (T = 5, N = 843).

In order to check the reliability of these different data, I first ignore the unobservable county effects and replicate the cross-sectional analysis done above by calculating simple OLS estimates using the pooled data. The results from this OLS are presented in column (1) of Tables 2.3-2.5 and are very similar to the cross-sectional results, suggesting that the use of the IPUMS data to impute the missing years of education will likely produce reliable results.[63] Further, as the results are largely similar across the three different possible panel data sets, these results suggest that estimation using these different panels should produce similar results throughout this analysis.

[62] IPUMS data is not currently available for 1890. For 1890 I impute educational enrollment using the ICPSR 2896 data for this year as discussed above. A discussion of the implications of this imputation for the reliability of my results is discussed below.

[63] Estimates using only the IPUMS data for educational enrollment delivers similar results in general. As an example, if I calculate the pooled OLS estimate using only the IMPUS data available for 1860, 1870, 1880 and 1900, the magnitude, sign and significance of all but one the estimates remains the same (the only exception is the urban variable which changes signs - though it is effectively zero in both cases). One important exception to this apparent reliability of the IMPUS data is the data available for 1850 which seems to produce unreliable results.

Second, I estimate equation (2.3) using the fixed effects panel data estimator, which allows for arbitrary correlation between the unobservable county effect and the other explanatory variables.[64] These estimates, which are presented in the second column of Tables 2.3-2.5, provide weak evidence that Freemasonry had a positive impact on the development of education as the estimates using both of the longer panels remain positive, with one estimate significant at the 5% level. However, the estimated effect using the shorter panel with the larger geographical coverage returns a point estimate that is effectively zero.

The difference between the point estimates from OLS and FE analysis may be attributable to a variety of causes, which I will address in the following sections. First, it may be the case that the Freemasonry variable is endogenous, in which case the first-differencing of the FE estimator will exacerbate the bias due to this endogeneity. Second, it is likely that educational enrollment is autoregressive as the availability of schools and teachers is likely to be correlated over time. In this latter case the strict exogeneity condition is violated and the estimates of (2.3) will be inconsistent. As this inconsistency is particularly severe for small T, this could explain the unusual results obtained above using the short panel Wooldridge (2002).

A major concern with the estimation of (2.1) and (2.3) is that there may be correlation between $X_{it}$ and $\epsilon_{it}$ due to the omission of a variable that affects both $X_{it}$ and $Y_{it}$. For Freemasonry, in particular, this may be a significant concern as the presence of strong community leaders is likely to affect the presence of Freemasonry and schools.

In the presence of this endogeneity, estimates of both equation (2.1) and (2.3) will be biased. For (2.1) this bias is given by

$$\begin{aligned}\hat{\beta} &= \frac{Cov(X_{it}, \beta X_{it} + \epsilon_{it})}{Var(X_{it})} \\ &= \beta + \frac{\sigma_{X,\epsilon}}{\sigma_X^2} \qquad (2.4)\end{aligned}$$

and for (2.3), since estimation is done by taking first differences and then using

[64] This is essential as it is likely that $Cov[X_{it}, c_i] \neq 0$ for most, if not all, of the covariates in $X_{it}$.

OLS, this bias is given as

$$\hat{\beta} = \frac{Cov(\Delta X_{it}, \beta\Delta X_{it} + \Delta\epsilon_{it})}{Var(\Delta X_{it})}$$
$$= \beta + \frac{Cov(X_{it} - X_{i,t-1}, \epsilon_{it} - \epsilon_{i,t-1})}{Var(X_{it} - X_{i,t-1})}$$
$$= \beta + \frac{\sigma_{X,\epsilon}}{\sigma_X^2(1-\rho_X)} \quad (2.5)$$

where $\sigma_{X,\epsilon}$ is the covariance between $X_{it}$ and $\epsilon_{it}$, $\sigma_X^2$ is the variance of $X_{it}$ and $\rho_X$ is the correlation of $X_{it}$ and $X_{i,t-1}$, all of which are assumed to be constant over time.

Correcting for the bias presented in equation (2.5) is possible if some other variable $Z_{it}$ is available to be used as in instrument, e.g. $Cov(\Delta Z_{it}, \Delta X_{it}) \neq 0$ and $Cov(\Delta Z_{it}, \Delta\epsilon_{it}) = 0$. Though no appropriate instrument is evident, as was also the case with the cross-sectional estimates presented above, the panel structure of the data allows us to study the bias is encountered in estimation of equation (2.3).

In particular, Anderson and Hsiao (1982) and Arellano and Bond (1991) propose using lagged levels (or differences) of the explanatory variables as instruments for estimation similar to this. In the case presented here, using a explanatory variable that was lagged two periods would indeed satisfy the exclusion restriction since $Cov(X_{i,t-2}, \epsilon_{i,t-1}) = 0$. However, for $X_{i,t-2}$ to be a good instrument, there must be a strong relationship between $Cov(X_{i,t-2}, \Delta X_{i,t})$. In my case, it is unlikely that there is a strong relationship between the level of Freemasonry in one decade and the change between two subsequent decades, especially since I am conditioning on a variety of socioeconomic variables. As a weak instrument, $X_{i,t-2}$ is likely to perform badly in practice.

Instead of instrumenting using $X_{i,t-2}$, I propose to estimate equation (2.3) using $X_{i,t-1}$. The first stage of this IV approach is guaranteed to be meaningful, and while the estimator will be biased, I can calculate the form of this bias. This bias is given by

$$\hat{\beta} = \frac{Cov(X_{i,t-1}, \beta\Delta X_{it} + \Delta\epsilon_{it})}{Cov(X_{i,t-1}, \Delta X_{it})}$$
$$= \beta + \frac{Cov(X_{i,t-1}, -\epsilon_{i,t-1})}{Cov(X_{it}, -X_{i,t-1})}$$
$$= \beta + \frac{\sigma_{X,\epsilon}}{\sigma_X^2} \quad (2.6)$$

Thus though I do not know the direction of the bias, if I know the sign of $\rho_X$, I can determine the direction of the bias by comparing these IV results to the OLS results. For Freemasonry, it is almost certainly the case that $\rho_X > 0$, that is that areas with more Freemasons in one period are likely to have more Freemasons in the subsequent period, as lodges require the presence of other Freemasons in an area to open and expand. Thus, after first-differencing, the IV estimates using $X_{i,t-1}$ should have lower bias than the OLS estimates.

I implement this IV estimation in two ways. First, I implement the pooled IV approach proposed by Anderson and Hsiao (1982). These authors propose using either $X_{i,t-1}$ and $\Delta X_{i,t-1}$ as instruments in a just-identified 2SLS estimation procedure. To check for the stability of the estimates, I consider estimation separately for each of the two possible instruments. Results for the estimate using the level are reported in column (3) of Tables 2.3-2.5 and the results for the difference are reported in column (4) of these tables.

Estimates obtained using the level, $X_{i,t-1}$, as an instrument provide evidence that the observed relationship between Freemasonry and enrollment is not being driven by endogeneity. As estimates using this approach are farther from 0 than the FE estimates in both Tables 2.3-2.4, which are the two panel data sets with a large enough T for this approach to be meaningful, this suggests that the bias in both equations is negative. Indeed, if $\rho_X > 0$, as postulated above, then the true relationship between Freemasonry and enrollment is actually larger than that estimated using either of these approaches. And though using the difference, $\Delta X_{i,t-1}$, as an instrument produces point estimates that are not different from zero, this is not necessarily evidence against a significant relationship between Freemasonry and enrollment but instead suggests that this difference is a weak instrument.

I then estimate (2.3) following the approach of Arellano and Bond (1991). This approach, which is recommended by Wooldridge (2002), uses all possible lagged values of $X_{i,t}$ as instruments. As my panels are relatively short, the number of instruments is actually quite limited - e.g. the panel that includes data from 1870-1900 has only six possible instruments. Column (5) of Tables 2.3-2.5 report the results from GMM estimation using these instruments.[65]

The Arellano and Bond (1991) approach delivers similar results to that obtained using the Anderson and Hsiao (1982) approach using levels. This can be seen by comparing columns (3) and (5) in these tables. That these two estimators gives similar results in the two longer panels and that estimates using both of these approaches are further from zero than the FE effect estimate is strong evidence that this relationship is not being driven by endogeneity.

[65] Pooled 2SLS using these instruments delivers similar results.

## 2.7 Exploring the Mechanism of Freemasonry's Impact

In the above section I documented the positive impact that Freemasonry had on the development of the education in the American West during the nineteenth. In this section, I examine the mechanism behind this relationship.

I begin by examining the relationship between ethnic heterogeneity, Freemasonry and educational outcomes. In this sub-section, I find that, in addition to observing that the share of the population that is native-born has a positive impact on enrollment, that the positive impact of Freemasonry is concentrated in more heterogeneous areas. Turning to panel data I find that this latter effect, that Freemasonry's benefit is concentrated in more heterogeneous areas, is robust to the inclusion of fixed effects; however, the population that is native-born is no longer found to have a significant effect. Finally, I consider several richer measures of heterogeneity which provide further evidence that Freemasonry's impact is concentrated in heterogeneous areas. In this final analysis I also find evidence to suggest that the observed positive correlation between share native-born and education is spurious.

In a second sub-section, I examine the relationship between religious heterogeneity, Freemasonry and educational outcomes. Here I find further (weak) evidence that Freemasonry's benefit is concentrated in more heterogeneous areas. In addition, using a panel data set on the size of the presence of different religious orders, I provide some weak evidence that the importance of religion in a county, as measured by the number of available seats in churches, had a positive effect on educational outcomes. However, I also find that the concentration Catholicism and Judaism had a negative impact on educational outcomes.

In a final sub-section, I explore a plausible mechanism for how Freemasonry affected educational outcomes. In particular I look at the relationship between Freemasonry and educational funding by looking at (1) public education income from taxation and public funds in 1850 and (2) local, county and state taxes in 1870. Here I find that evidence that areas with more Freemasons had higher local taxes as a share of total taxes, consistent with the hypothesis that areas with more Freemasons were able to overcome the common good problem. Interestingly, I also find evidence that areas with more Freemasons in 1850 had a higher share of teachers per student in private schools. This suggests that Freemasons have a stronger preference for education than other citizens. While my capture does not capture this level of heterogeneity, it is consistent with my claim that Freemasons valued education more than others.

### 2.7.1 Freemasonry and Ethnic Heterogeneity

The work of Goldin and Katz highlighted the negative relationship between ethnic heterogeneity and educational development in American history. Interestingly, while I consistently find a positive relationship between educational enrollment and the share of a population that is of native parentage, which supports the claim of these authors, this result does not seem to be robust to the inclusion of fixed effects. Indeed, while in Table 2.5 I show that this measure of homogeneity loses its statistical significance when controlling for unobserved heterogeneity, estimation using panel data that spans from 1860 to 1900 produce significants *negative* estimates of the relationship between *homogeneity* and educational outcomes as shown in Table 2.4.

In this sub-section, I examine both this empirical phenomenon as well as the relationship between Freemasonry and ethnic heterogeneity. As Freemasonry similarly shows (see Section 2.8.2) a strong positive relationship with ethnic homogeneity, as measured by the share of the population of native parentage, in pooled OLS regressions and a negative relationship in fixed effects panel analysis, this suggests that the impact of Freemasonry and ethnic homogeneity/heterogeneity may be related. And as Freemasonry encouraged the expansion of public education by bringing men of all types together in support of a common cause, I might expect that the Freemasonry effect would be more pronounced in more heterogeneous areas.

My analytical strategy for looking at the relationship between Freemasonry, ethnic heterogeneity and educational outcomes is very straightforward. In particular, I simply augment my previous analysis of educational enrollment to include an interaction term between Freemasons per capita and the share of the population of native parentage. In Table 2.6, where I report the results using the cross-sectional data available for 1870, I find that this interaction term has a very significant negative coefficient. As areas that have a larger value of "% native" are more homogenous, this coefficient indicates that the impact of Freemasonry on education is attenuated in areas that are more socially homogenous. Indeed, the impact of Freemasonry is four-fold greater in a community with no individuals of native parentage as compared to a community that is 100% native.

Table 2.7 reports the results from this analysis using the 1860-1900 and 1870-1900 panel data sets that contain data on the share of the population of native parentage. Note that here, unlike before, I do not include the results that correct for endogeneity as the results do not change substantively.

The central result of this table is that Freemasonry's impact on education seems to be the largest, by far, in areas that are the most heterogeneous. Indeed, in areas that are very homogenous - e.g. the share of the population of native parentage is close to 1 - the impact of the number of Freemasons on educational outcomes is estimated to be quite small.[66] However, in areas that are very heterogeneous, an increase in the number of Freemasons is predicted in a county is expected to have a very strong impact on education: if the number of Freemasons per capita increased by 0.01 - e.g. one additional Freemason for each one hundred local residents - in an area where 50% of the population was of native parentage, this would lead to a 6-7% point increase in the educational enrollment rate. Perhaps most interestingly is that this result is one of the few empirical results that is strongly significant, with the same sign, across both data sets and specifications.

One potential concern with this analysis is that the measure of homogeneity used thus far, the share of the population of native birth, does not truly reflect the homogeneity of a community. Indeed, it is possible that a community with few individuals of native birth can actually be quite homogenous. Thus, a second possible interpretation of the result for this interaction term is that the Freemasonry effect is concentrated in areas with a large number of recent immigrants.

In order to explore this issue, I restrict my analysis to 1870, 1890 and 1900 which allows me to augment my analysis in two ways. First, I can divide the population native born into the population native born of native parentage and the population native born of foreign parentage. Second I can construct a measure of the ethnic fractionalization of an area. This measure, which I call the ethnic fractionalization index (EFI), is constructed as

$$EFI = 1 - \sum_{i=1}^{n} \frac{P_i}{Total\ Population} \tag{2.7}$$

where $P_i$ is the population in group $i$ and $i$ indexes the different ethnic groups distinguishable in the census. While these censuses allow me to identify the origin of birth of foreign born individuals, in practice this only accounts for 15% of the total individuals included in my data. Thus, much of the variation in EFI is driven by the ratio of the number of native born individuals with native-born parents as compared to those with foreign-born parents. The data section above has more details on this measure.

Table 2.8 reports the analysis using these two more disaggregated measures of ethnic homogeneity. Columns (1)-(7) provide estimates using pooled OLS and columns (8)-(14) provide the fixed effects panel estimation results.[67] Analysis examining the two measures of parent nativity and EFI are done separately as these variables are closely related and it is difficult to separately identify the impact of each.

[66] This is calculated by multiplying the coefficient for the interaction term times 1 - or close to 1 - and adding this to the coefficient for the Freemason term by itself.

[67] Estimates using the endogeneity correction are omitted as they perform quite poorly in this short panel.

In comparing column (1), which repeats the analysis used above, with column (2), which disaggregates the share native by type of parentage, an interesting phenomenon emerges: the strong positive relationship between nativity and educational outcomes disappears. This is verified by columns (8) and (9) which report the same results controlling for unobserved county-level heterogeneity. This suggests that the observed positive relationship between nativity and educational outcomes is perhaps spurious.

In columns (3) and (10) I include my Freemasonry variable. Similar to the analysis above using 1870-1900 data in Table 2.7 I find a significantly positive point estimate for the pooled OLS and no effect for the fixed effects estimation. However, when I augment this analysis to include an interaction term between Freemasonry and the two measures of homogeneity, in columns (4) and (11) I find a positive, large and significant coefficient for Freemasonry.

Interestingly, the estimated coefficient on both of the interaction term is negative, which is the same as I found for the aggregate share of native-born in Table 2.7. Thus, while it does not seem that nativity effects educational outcomes directly, it is the case that the benefit of Freemasonry is concentrated in areas that are more heterogeneous.

Columns (5) and (12) consider the impact of my ethnic fractionalization index (EFI) on educational outcomes. The effect of EFI is surprisingly weak, as an area with an EFI of one - where every individual would from a different county - is expected to have an enrollment rate that is only four points lower. However, the inclusion of an interaction term with Freemasonry leads to much larger and significant estimates of its impacts. In particular, it seems that this ethnic heterogeneity has a *positive* impact on educational outcomes in the presence of Freemasonry. This is further evidence that the benefit of Freemasonry is concentrated in more heterogeneous areas.

### 2.7.2 Freemasonry and Religious Heterogeneity

The above analysis of educational outcomes focuses on the estimation of the importance of Freemasonry in the development of education as compared to a variety of socioeconomic variables. In this sub-section, I augment this above analysis to include a variety of religious variables. An important limitation of this analysis is that data on religion is only available for four decades: 1850, 1860, 1870 and 1890.

In Table 2.9 I report the results from this analysis. Columns (1), (6) and (11) use the same parameterizations that were utilized throughout Section 2.6.2 using this more restricted, and non-sequential, panel data. Column (1) reports the simple pooled OLS results, column (6) reports the results using fixed effects panel estimation and column (11) reports the results using the estimator of Arellano and

Bond (1991) to correct for endogeneity.[68] These results are largely the same as those found before, as the Freemasonry effect remains robust after correcting for endogeneity, agricultural areas seem to have higher educational enrollment rates and the share of men has a negative impact on education.

Columns (2), (7) and (12) repeat the same analysis including four measures of the presence of religion within a county: the number of Catholic pew seats per capita, the number of seats in Jewish churches per capita, the number of total seats in all churches per capita. As data on the number of Catholic pew seats is not available in 1870, columns (3), (8) and (13) report the results without this measure of local Catholicism, and then columns (4), (9) and (14) add the religious fractionalization index (RFI), my measure of religious heterogeneity which is discussed in the data section, to this analysis. Finally, columns (5), (10) and (15) augment this analysis to include an interaction term between the RFI and the concentration of Freemasons in hopes of understanding if the relationship between Freemasonry and social homogeneity observed above is similar for religious heterogeneity. This analysis demonstrates several interesting phenomenon about the relationship between religion and educational outcomes in the nineteenth century Western United States.

First, the negative correlation between the importance of Catholicism in a county and education is consistent with the existing historical literature on the relationship between Catholicism and public education in the United States. While this effect does disappears when I adjust for unobserved county effects, this could largely be an artifact of the data as there are only three years of data available for the prevalence of concentration (1850, 1860 and 1890) and the final year is separated by a gap of thirty years.

Second, there does seem to be a positive relationship between religion and education. While the estimated relationship between religion and education is only significant at the 5% level in only two of the six specifications, it is positive in all specifications and significant at least the 10% level in all but one specification.

Third, the pooled OLS analysis suggests that areas with higher levels of religious heterogeneity have higher educational attainment. While the significance of this effect disappears in the fixed effects panel analysis, a positive coefficient does remain. This is further evidence that heterogeneity is not necessarily bad for the provision of public goods.

Finally, the interaction term between Freemasons and the RFI provides further evidence that the benefit of Freemasonry was concentrated in heterogeneous areas. Indeed, the point estimate for this interaction term is weakly positive - at

[68] I only report the results from this estimator as it seemed to give the most stable results above and does so here as well. The results from the two other estimators considered above are very similar for this analysis though the Anderson and Hsiao (1982) estimator using differences still seems to suffer from a weak instrument problem.

the 10% level - in both the pooled OLS results as well as the panel results that are corrected for endogeneity (following Arellano and Bond (1991)).[69]

### 2.7.3 Freemasonry and Taxation

The hypothesized mechanism underlying the relationship between Freemasonry and the provision of education within a locality is that the presence of Freemasonry helped communities overcome the common good problem. In this sub-section, I use taxation data to examine an important prediction of this hypothesis. In particular, as schools were funded and provided at primarily the village/city level, my hypothesis would predict that villages/cities with more Freemasons would pay higher levels of local taxes, all other things equal.

I separate the analysis of taxation data into two sections as the taxation data that is available is in two different forms. The first of these sections focuses on data from 1850 U.S. Census which has data on the amount of funding for public schools from taxation and public funds. Additionally, I look at the impact of Freemasonry on the number of public versus private school teachers, which I argue is an indirect measure of the amount of wealth being directed towards these two different types of schools. In this period, which corresponds to the time during which the influence of Freemasonry was just beginning to expand, I find no effect of Freemasonry on public funds or taxation. Additionally I find that areas with more Freemasons were more likely to have fewer public teachers per capita and more private teachers per capita, a result which is consistent with the result I found above.

Second I use data available for 1870 on the amount of local, county and state taxes paid by each county. In this data, which corresponds to the period following the rapid expansion of both the geographical spread as well as the concentration of Freemasonry, I find evidence that regions with higher concentrations of Freemasons pay higher taxes of all kind, after inclusion of a variety of controls for income, wealth and social homogeneity. Additionally, I find that areas with more Freemasonry paid more local taxes as a share of total taxes than other counties. As education funding was typically done at the village or local level for the Western United States, this suggests that areas with more Freemasons where indeed more able to overcome the common good problem and increase local taxation.

Unfortunately, these data do not allow me to connect the relationship between local taxation in 1870 and the amount of revenue that public schools are receiving from taxes. Thus in addition to being unable to show that Freemasonry has a causal impact on local taxation, I am not even able to draw a direct connect between this local taxation and the amount of the taxation that is going towards

[69] Both the concentration of Freemasons as well as the interaction term were treated as potentially endogenous variables for this analysis.

public schools. However, as public education was typically the largest recipient of local taxation, this evidence is consistent with my hypothesis.

In Table 2.10 I examine the relationship of Freemasonry with both the revenue and expenditures of public and private schools. Consistent with my examination of 1850 public school enrollment data, I find no evidence of an impact of Freemasonry on revenue. Indeed, the point estimates on Freemasonry in columns (6) and (8) suggest that Freemasonry did not have much of an impact on . This is consistent with the evidence presented above that Freemasons where not more likely to migrate to areas with an *ex ante* preference for public education and that their influence was still quite limited at this point.

However, I find evidence that suggests that expenditures by public schools were actually lower in areas with more Freemasons. Though the census data does not include a measure of expenditures by these schools, I use the number of public school teachers per potential student as a proxy for expenditures. And as shown in column (2) of Table 2.10, the number of public school teachers actually seems to be significantly lower in areas with more Freemasons. The point estimate on Freemasonry found in column (2), -0.20, is actually quite large as it indicates that a one standard deviation increase in the number of Freemasons would reduce the the number of public school teachers by one-tenth of a standard deviation.

Interestingly I find evidence that areas with more Freemasons are actually likely to have higher expenditures on private academies. This is consistent with the evidence I found above on private school enrollment and indicates that Freemasons either (1) had a strong preference for education for their children and were willing and able to pay for or (2) were more likely to migrate and live in areas with more private schools.

I investigate the relationship between Freemasonry and taxation in 1870 in two different ways. First, using the same parameterization used above for education, I look at the relationship between Freemasonry and the level of taxation at different levels of geographic aggregation. Second, my hypothesis predicts that the impact of Freemasonry should be largest for local taxes. Thus I compare the impact of Freemasonry on the ratio of local, either village or county, to state taxes. Focusing on this ratio is also particularly helpful because state taxes are likely to be very strongly correlated with local wealth.

In Table 2.11 I examine the relationship between Freemasonry and the level of local (village), county and state taxation. In all cases, I find a positive, though not always significant, relationship between the concentration of Freemasons and the level of taxation despite the inclusion of a variety of controls for both local income and wealth as well as measures of social homogeneity. Interestingly, while the measures of local, agricultural and manufacturing, have a positive relationship with taxation, income seems to have a negative effect when it has an effect at all. Consistent with the model proposed by Goldin and Katz (1999), social homogeneity as well as the number of people living in small towns have a positive impact on county taxes. However, social homogeneity does not seem to effect either local or state taxes.

Though the relationship between Freemasonry and taxes is highly significant (1% level) for both county and state taxes, the relationship is much weaker for local taxes. Though there is a significant overall impact at both the county and state level the effect seems to be largely driven by the 10% of counties that derive at least two-thirds of their income from manufacturing, though the relationship is still weakly positive in the more agricultural regions. The weak effect observed for local taxes may be at least party driven by the fact that nearly half of the localities reported having no local taxes. This downward censoring may exclude many counties at an early stage of development where the Freemasonry impact would potentially be particularly significant.

The evidence for the relationship between Freemason concentration and the ratio of local to state taxes is a bit more mixed as is shown in Table 2.12. Indeed, while the relationship seems to be positive for the ratio of village/city taxes to state taxes, there seems to be a negative - though weak - impact on the county to state tax ratio. Interestingly, only one of the two social capital measures suggested by Goldin and Katz has a significant effect and is similar to the Freemasonry effect as it is only significant at the village level. Though the Freemason variable only has a significantly positive effect among the manufacturing villages, the effect among the more agricultural villages is still positive.

## 2.8 Robustness Checks

An important concern with the results presented above is that Freemasons are either (1) migrating to areas that have higher provision of public education or (2) more likely to form in areas with public education. In this section, I provide two types of robustness checks that suggest that reverse causality is not driving the result that I observe.

I first examine the relationship between Freemasonry and education in the 1850. As this early period corresponds to the very beginning of the spread of Freemasonry in the West, and when the influence of Freemasonry was still quite weak, it allows me to explore whether Freemasons were indeed more likely to migrate and form in areas with better public education. Though I do find a weak positive relationship between Freemasonry and private education during this period among counties with very high concentrations of Freemasonry, this is evidence of Freemasons' appreciation for education rather than the potential for a community to a develop public education system (as suggested in Section 2.7).

I then explore the spread of Freemasonry in the West in more detail. This allows me to examine the possibility that Freemasonry is simply functioning as a proxy for some other economic variable. I demonstrate that the development of Freemasonry was not decisively affected by any observable variables. As the spread of Freemasonry does not seem to have been affected by important economic variables such as income and wealth, this is evidence that Freemasonry is unlikely to be functioning as a proxy for some other variable.

### 2.8.1 Education in the Mid-19th Century

Both Freemasonry and educational enrollment data are relatively limited in the Western United States in the mid-19th century. However, understanding the relationship between Freemasonry and education in 1850 is an essential part of this analysis as Freemasonry was just beginning to re-emerge in this year. Thus, if we observed a significant positive relationship between Freemasonry and education in 1850, this would be strong evidence that Freemasons were likely to migrate to areas with existing educational systems or that Freemasons themselves had a positive impact independent of the organization.

However, there does not seem to a significant positive relationship between Freemasons and total enrollment in 1850. This is demonstrated in column (2) of Table 2.16. Both columns (1) and (2), which consider total enrollment within counties, indicate that income and wealth have a positive impact and the one measure of social homogeneity that is available, the share of the population in small towns, has a weak positive effect. Interestingly, a large adult male population seems to be associated with low educational attainment, likely reflecting the many frontier counties in these data.

These data also allow me to separately analyze the impact of Freemasonry on public and private school enrollment. And while there is no significant relationship between Freemasonry and enrollment in public schools, areas with higher concentration of Freemasons have significantly higher enrollment in private academies. As academies only account for 3% of total enrollment during this period, this affect is muted when the date are pooled. The estimated coefficient of $\sim$2 implies that for each additional Freemason that there are two additional children in private academies. Though it is impossible to attach causality to these estimates, this evidence is consistent with the strong preference among Freemasons for education.

An important caveat to these results is that this analysis only considers data from Arkansas, Illinois, Indiana, Missouri and Wisconsin. Though California, Iowa and Oregon all have data available for this period, I excluded them from this analysis as most the counties in these states were thinly populated. And though both Louisiana and Texas had large, geographically distributed populations, I do not have Freemason data for these states in 1850 as of yet.[70]

### 2.8.2 The Spread of Freemasonry

Freemasonry spread rapidly throughout the states of the Western United States during the second half of the nineteenth century in terms of both size and geographical coverage as shown in Figure 2.4. From a relatively small presence of only 0.2% of the population in 1850, Freemasons accounted for over 1% of the population by 1870 - a share that was maintained through 1900. The geographic spread of Freemasonry was equally impressive: while less than half of the counties in 1850 had Freemasons, by 1900 Freemasons were present in all but 1% of the counties of the Western United States. Though Freemasons were found throughout the country, their concentration varied massively, from a low of 0.05% of the population to upwards of 2 or even 5% of the population.

What explains the variation in the presence of Freemasons in a locality during the nineteenth century? And what explains its rapid growth in the Western during this period United States? In this section, I look at the possible influence of socioeconomic factors on the spread of Freemasonry.

First, I explore what characteristics were correlated with a larger prevalence of Freemasonry and find that areas that had a larger religious presence and were more homogenous had more Freemasons. Interestingly I find that the share of Catholics was negatively related to Freemasonry, as suggested by the historical literature, while the relationship with Judaism was positive during the mid-19th century and became negative by the end of the century. Finally, areas that were more religiously heterogeneous, as measured by a religious fractionalization index that I construct here, had higher concentrations of Freemasons.

Second, as I am interested in understanding the determinants of Freemasonry, in addition to simply the characteristics that are correlated with Freemasonry, I turn to panel data that will allow me to control for unobserved county heterogeneity. After controlling for this potential heterogeneity, ethnic homogeneity has a negative effect on the presence of Freemasonry, the impact of religious heterogeneity disappears and for the first time a measure of wealth - in this case, agricultural wealth - enters significantly. These latter results are robust to correction for possible autocorrelation.

While the historical literature has proposed a variety of factors that may have been correlated with the presence of Freemasonry, the lack of comprehensive quantitative data in previous studies has made it difficult to test the relative importance of different factors. In Table 2.13 I examine the relationship of a variety of potentially important socioeconomic variables with Freemasonry. This

[70] These data were not available in the Iowa Masonic Library though I hope to obtain them in the future.

table reports conditional correlations, which are calculated by regressing the concentration of Freemasons against a variety of socioeconomic variables, which are disaggregated by year so that I can examine the potentially changing character of Freemasonry over time.

In columns (1)-(6) of Table 2.13 I focus on the relationship between Freemasonry and several potentially important economic and demographic characteristics in 1850-1900. The only effect that I find consistently across nearly all years in this analysis is that counties with a larger population in agricultural areas, measured by either the share of population living in urban areas (the opposite), the amount of agricultural capital or the amount of agricultural output, are more likely to have more Freemasons. Suprisingly, despite the high cost of joining the organization, which would suggest that wealthier areas would be likely to have more Freemasons, I do not find evidence that total income or wealth are related to the presence of Freemasonry.

The results from the first six columns of this table table also provide evidence that the nature of the Freemasonry organization changed toward the end of the nineteenth century as suggested by Tabbert (2005) and others. The most striking piece of evidence that Freemasonry was evolving from an exclusive organization into a social club between the late 1880s and the turn of the century is that the importance of the share of men as a correlate of Freemasonry changes dramatically between 1890 and 1900. Though one might expect that Freemasons would be more prevalent in areas with more men, as there would be more potential candidates, in practice the family is a very important of being a Freemason. Thus, while the relationship between Freemasonry and the share of men is weakly positive and stable from 1850-1890, the large increase in both the point estimate and significance of this variable is consistent with the evolution of Freemasonry into a 'club for men'.

The relationship between Freemasonry and religion is explored in columns (7)-(10) of Table 2.13. While my empirical approach does not allow me to identify the faiths of Freemasons themselves, I can examine the religious composition of the communities where Freemasonry is most prevalent. In all years for which there is religious I find that counties with a larger religious presence, as measured by the total number of seats in all churches, have more Freemasons as a share of the local population. This likely reflects either the complementarity of these social institutions or the importance of religion to Freemasons.

The antagonistic relationship between the Catholic Church and Freemasonry is supported by the negative relationship found between the importance of a Catholic Church in an area and the importance of Freemasonry. Interestingly, while a similar relationship was found during the later part of this period for the Jewish faith, a positive relationship is observed between the importance of Judaism and Freemasonry in 1850. This result, which is robust to the exclusion of communities with the highest concentrations of Jewish seats per capita, indicates that Freemasonry spread particularly in areas that were Jewish in an early period. While this could reflect a relationship between Judaism and Freemasonry, it more likely reflects the urban nature of Freemasonry as the concentration of Jews in an area is a better measure of urbanization than the share of the population in cities with over 25,000 people.

Interestingly, religious fractionalization, which is a measure of the religious heterogeneity of a county as discussed in the data section above, seems to have a very strong *positive* relationship with Freemasonry during 1850-1870. This is perhaps an unusual finding. However, it is consistent with the fact that Freemasonry's impact seems to be more important in heterogeneous areas.

This positive relationship between Freemasonry and religious fractionalization disappears in 1890. I also find that the share of the population of native parentage, a share of ethnic homogeneity, increases during this period. Though this analysis is not sufficient to make any strong claims, this does give provide some evidence about the nature of the transition that the Freemasonry organization experienced in the late nineteenth century. In particular, as the organization lost its elite character and became more inclusive it also became more nativist and homogenous.

In the above sub-section, I examine the correlation between a variety of socioeconomic characteristics and Freemasonry. While these correlations are interesting in their own right, as they provide valuable information about the type of communities where Freemasonry was more prevalent, these previous estimates should not be interpreted as measures of the impact of these different factors on the presence of Freemasonry. In particular it is likely that there are unobservable variables, such as the amount of 'social capital' in a community, that is likely to influence both the concentration of Freemasonry and some of the explanatory variables. In this section, I will use a panel data approach to compensate for the potentially confounding effect of unobserved county heterogeneity.

The approach is straightforward. The share of the population that is Freemason in county $i$ and at time $t$, $F_{it}$, is modeled as

$$F_{it} = X_{it}'\beta + \omega_i + \epsilon_{it} \tag{2.8}$$

where $X_{it}$ is a variety of socioeconomic variables that are likely to affect Freemasonry, $\omega_i$ is a fixed effect that is possibly correlated with $X_{it}$ and $\epsilon_{it}$ is a stochastic shock. Equation (2.8) is then estimated using three different sets of panel data. The first of these data sets has the most coverage in terms of time, spanning from 1850-1900, but contains the least geographic coverage as it uses the earliest available county definitions which tend to be more aggregated than later counties (T = 6, N = 451). The third data set covers the fewest years, only 1870-1900, but offers the greatest geographical variation (T = 4, N = 1034) while the second, which covers 1860-1900, is a balance between geographical variation and years covered (T = 5, N = 843).

Table 2.14 reports the results from the simplest estimate of equation (2.8) which are broadly consistent with the results found above. Columns (1)-(3) report standard OLS estimates using the data constructed for the panel estimation, which are calculated to demonstrate the comparability of these data will the previous data, while columns (4)-(6) report results from fixed effects panel estimation. These fixed effects estimates show that the importance of agricultural in an area has a positive impact on the development of Freemasonry, which was suggested in the above analysis using conditional correlations. However, while the above correlational analysis indicated that communities with higher ethnic homogeneity had higher levels of Freemasonry, especially in a later period, the results from columns (4)-(6) suggest that this homogeneity has a negative impact, if any, on the share of Freemasons.

Columns (7) and (8) of Table 2.14 also provide weak evidence that the point estimates for the religious variables in the conditional correlations reported in the previous sub-section may have a causal interpretation. While the point estimates on the Freemason variables are never significant, identification is difficult as the level of religiosity and its heterogeneity does not vary significantly over the 40 year horizon that is covered by these data.

A contention that may be raised with the estimation of equation (2.8) is that it ignores the possibility of autocorrelation in the Freemasonry variable. Indeed, as lodges require the presence of other Freemasons in an area to open and expand it is likely that areas with more Freemasons in one period are likely to have more Freemasons in the subsequent period.

In order to both correct for the possible bias due to ignoring this autocorrelation, as well as to estimate the correlation of the concentration of Freemasonry across decades, I thus estimate

$$F_{it} = \alpha F_{i,t-1} + X_{it}'\beta + \omega_i + \epsilon_{it} \tag{2.9}$$

where $F_{i,t-1}$ is the lagged level of the concentration of Freemasons in a county and the rest of the notation is the same as before. As my panel data has only 4-6 periods of data, depending on the data set used, I face the classic "small T" problem and standard approaches for estimating the AR1 model are not available. Thus I estimate equation (2.9) using the approaches of Anderson and Hsiao (1982) and Arellano and Bond (1991) which are discussed in more detail in Section 2.6.2 below.

Table 2.15 reports the results from the analysis that adjusts for the possibility of autocorrelation in the dependent variable. Columns (1)-(6) report the results for each of the two estimators for the three possible panel data sets. In each I find that there is a significant degree of autocorrelation. While the point estimates vary from 0.10-0.50 across the various samples and specifications, the fact that the point estimate is always positive and significant (usually at the 1% level) indicates that the presence of Freemasonry in a county has an impact on Freemasonry in that county a decade later.[71]

Interestingly, once I correct for this this autocorrelation, the estimates of the impact of the concentration of adult males becomes very strong and significant. Indeed, I find evidence that communities that have a large concentration of adult males are less likely to have large Freemason communities. This result, which is significant across all specifications, indicates that the presence of families was a driving factor in the expansion of the organization in the West.[72]

A similar analysis for the panel data that includes religious variables does not produces any new significant results. However, these results, which are presented in columns (7)-(8) of Table 2.15, do demonstrate the robustness of the negative relationship between the share of males in a county and the prevalence of Freemasonry.

## 2.9 Conclusion

In this paper I explored the role that American Freemasonry played in the development of the educational system in the United States. Using a novel panel data set of American Freemasonry during the 19th century, I documented that counties with higher concentrations of Freemasons had higher primary school enrollment rates in the late 19th century. And the impact of Freemasonry was quite important: a one standard-deviation increase in the concentration of Freemasons was equivalent to a one-fourth of a standard-deviation increase in the primary school enrollment rate. The availability of panel data for both enrollment rates and Freemasonry allowed me to rule out both reverse causality and endogeneity as possible explanations for this observed effect.

And, in what is a striking example of path dependence, I showed that the concentration of Freemasonry in the 19th century also had a significant positive impact on secondary school enrollment in the beginning of the following century. This result is particular significant given that most reports indicate that the influence of Freemasonry had largely waned by the end of the 19th century.

The evidence presented here indicate that Freemasonry impacted the American education system by helping communities attenuate the common good prob-

[71] If I restrict the sample to only counties with non-zero Freemasons, I find a very similar effect.

[72] An identical result is found using data restricted to counties with non-zero Freemasons.

lem. This was demonstrated in two ways. First, the impact of Freemasonry was particularly meaningful in areas that typically suffer the most from common good problems, i.e. those that were the most heterogeneous ethnically and religiously. Second, counties with higher concentrations of Freemasons had higher levels of local taxation. This result, which is consistent with both the rhetoric of Freemasons and other historical studies of Freemasons, indicates the powerful development impact that a social institution with a relatively small number of individuals can have.

Figure 2.1: National Prominence of American Freemasonry

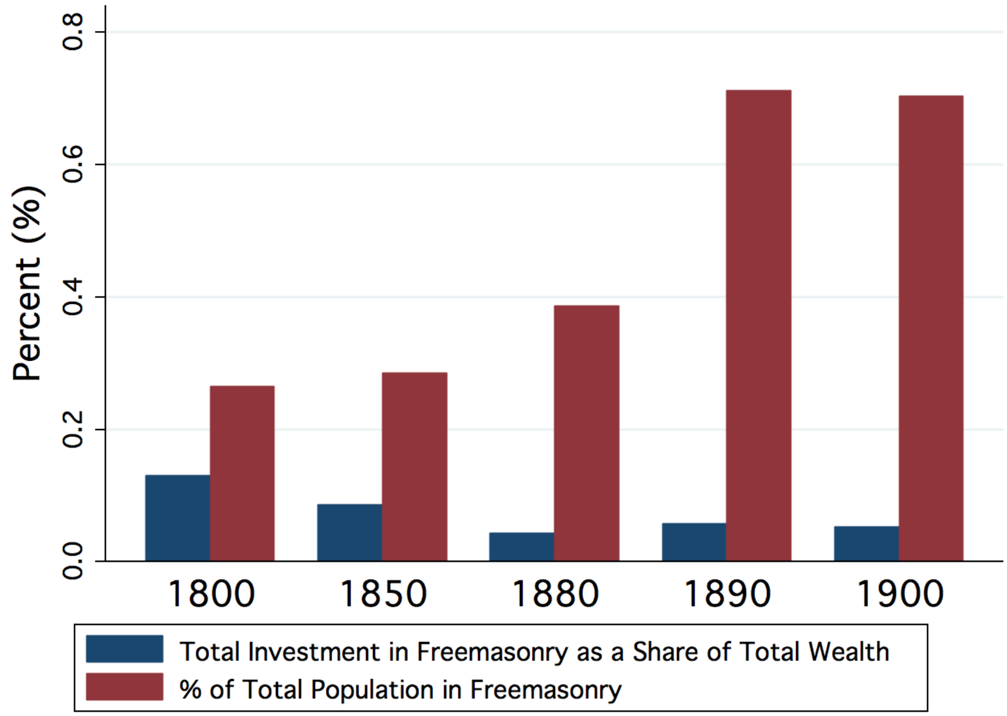


Figure 2.2: Freemason Capital Relative to Total Physical Capital
(state means and standard deviations)

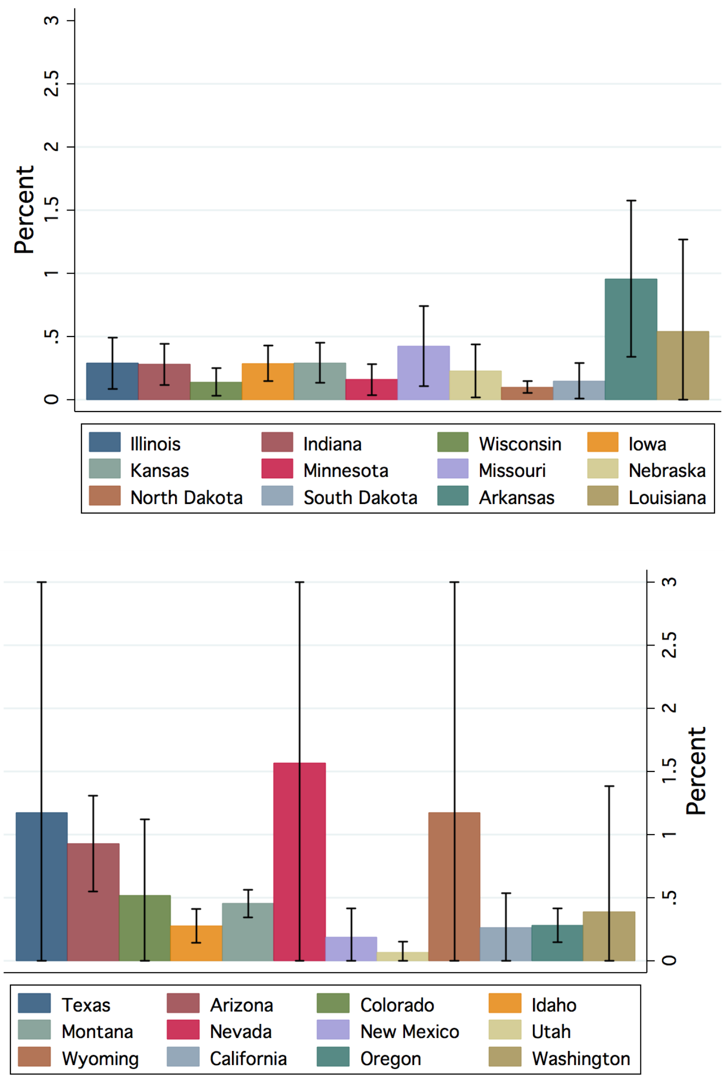
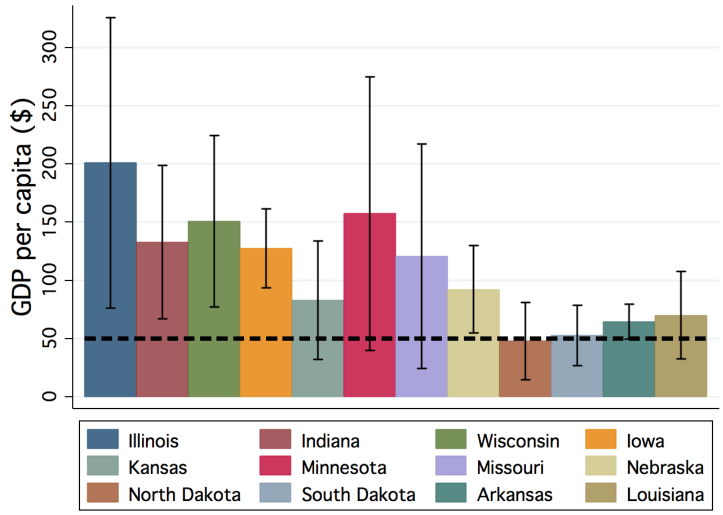


Figure 2.3: GDP per capita and the Cost of Freemasonry
(state means and standard deviations)



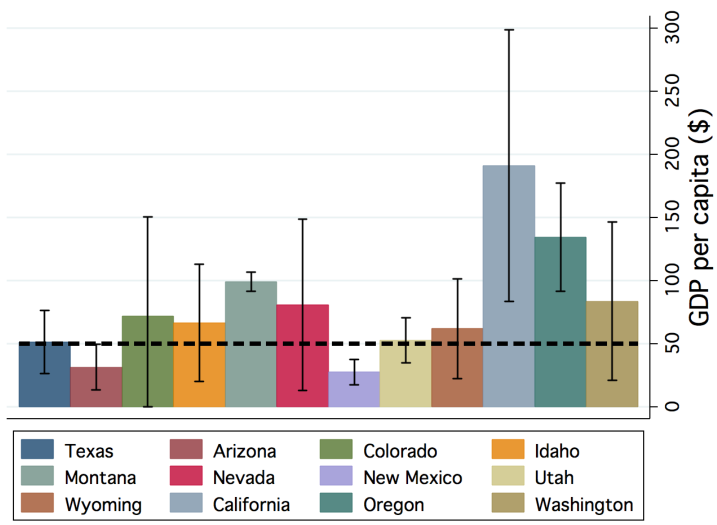



Figure 2.4: Spread of Freemasonry in the Western United States

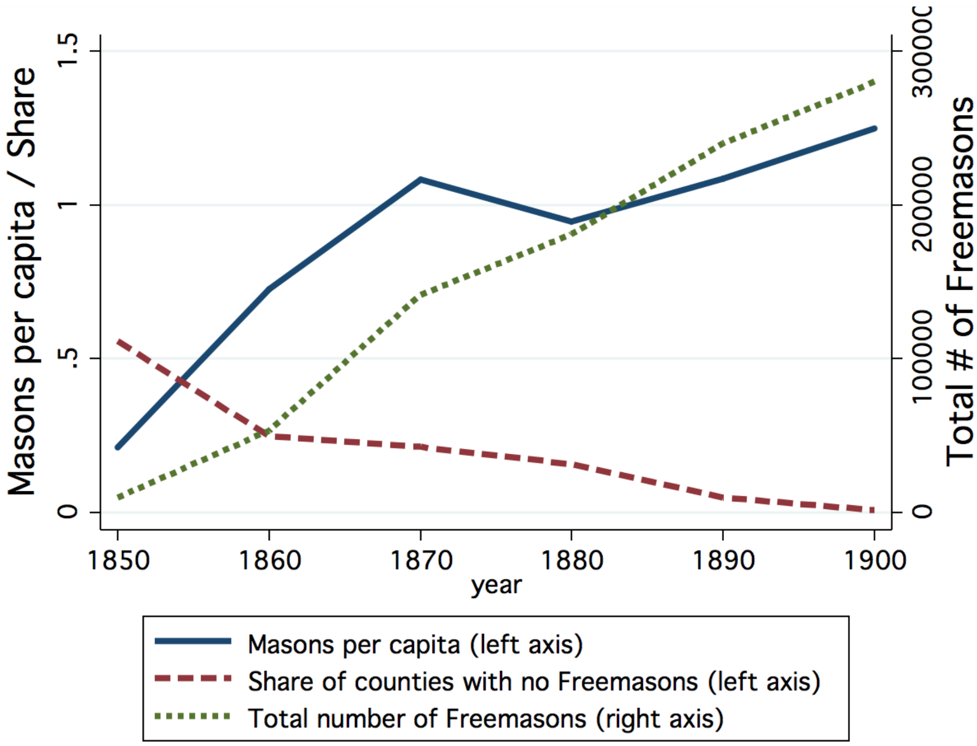


Figure 2.5: Freemasons in the Western United States: 1850

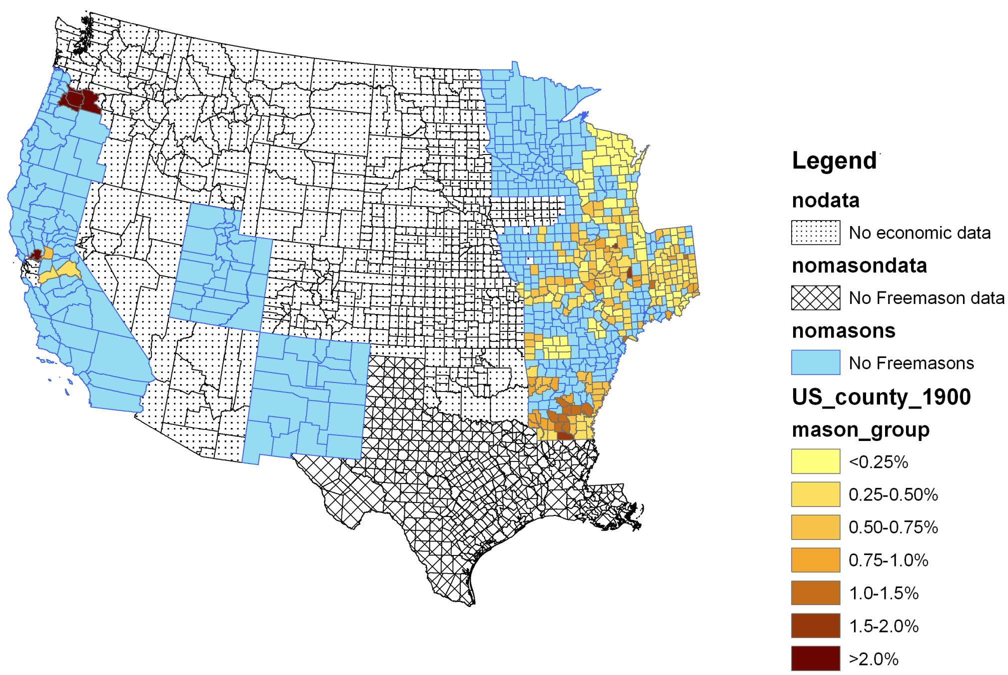


Figure 2.6: Freemasons in the Western United States: 1860

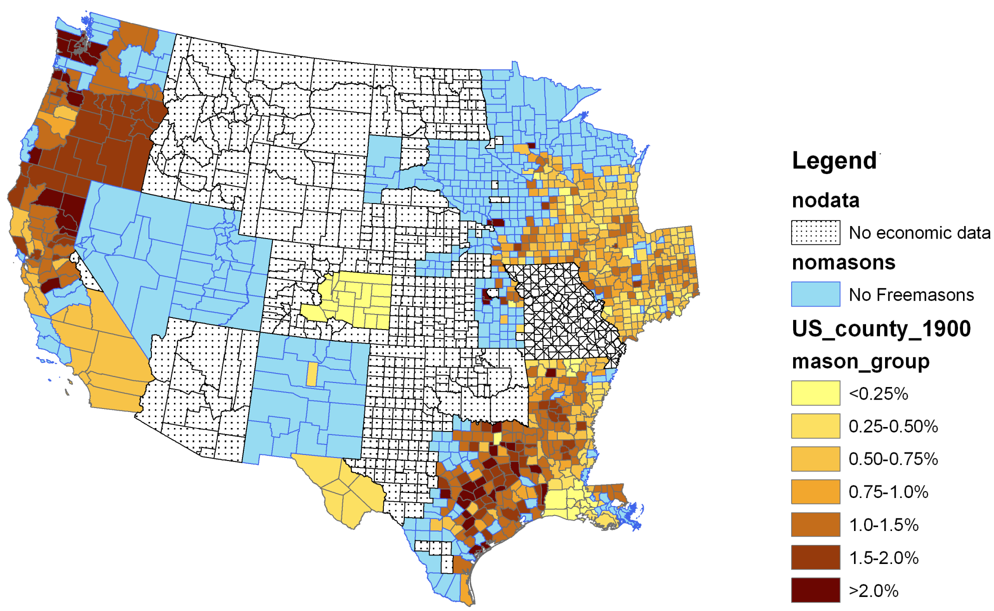


Figure 2.7: Freemasons in the Western United States: 1870

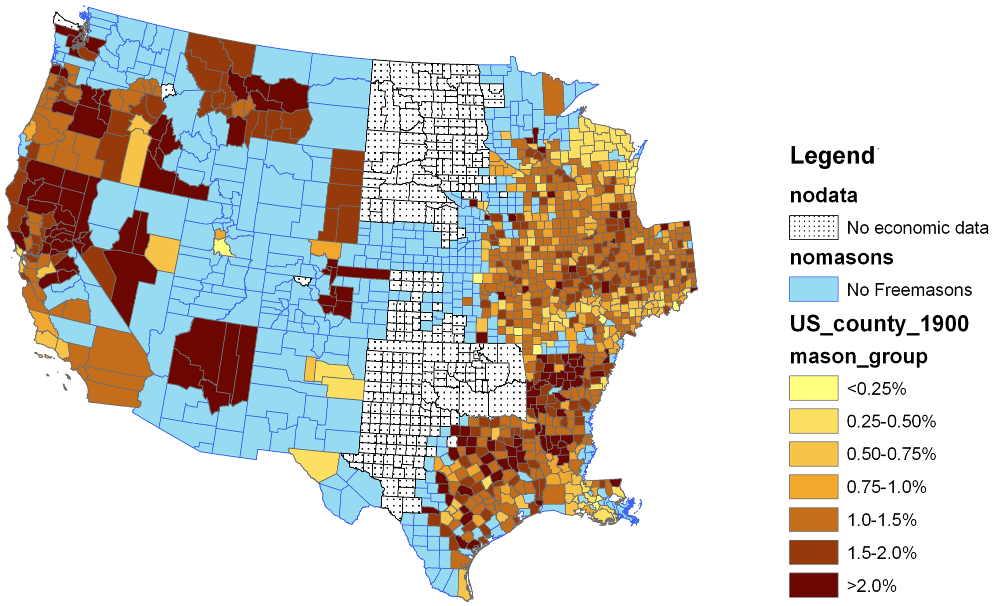


Figure 2.8: Freemasons in the Western United States: 1880

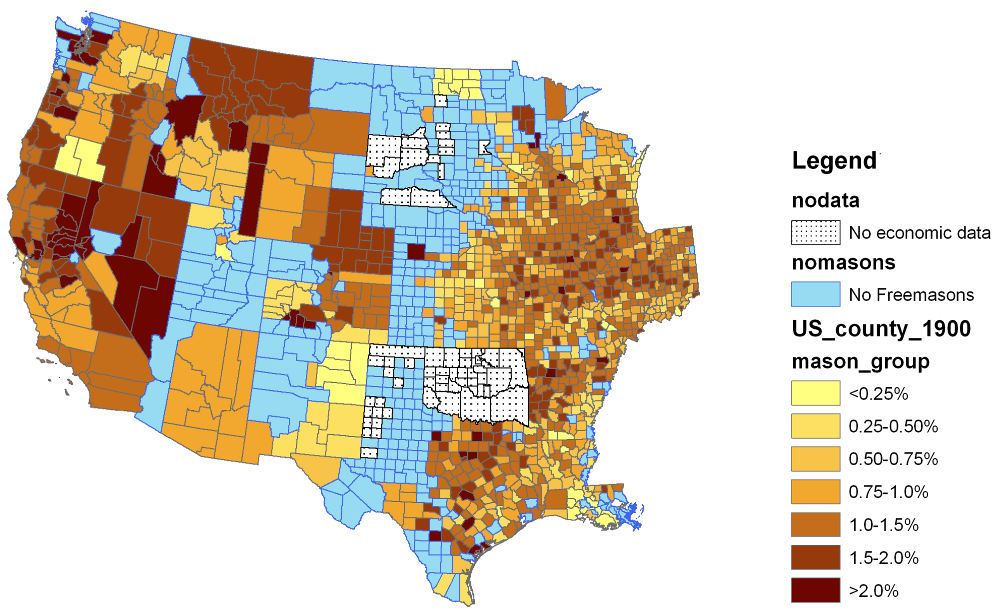


Figure 2.9: Freemasons in the Western United States: 1890

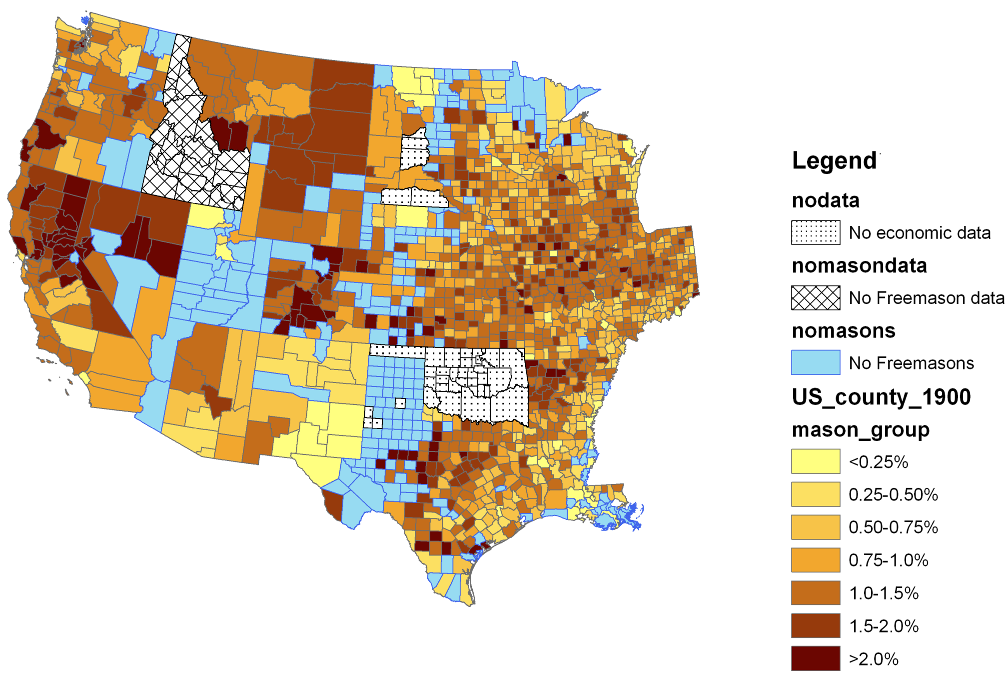


Figure 2.10: Freemasons in the Western United States: 1900

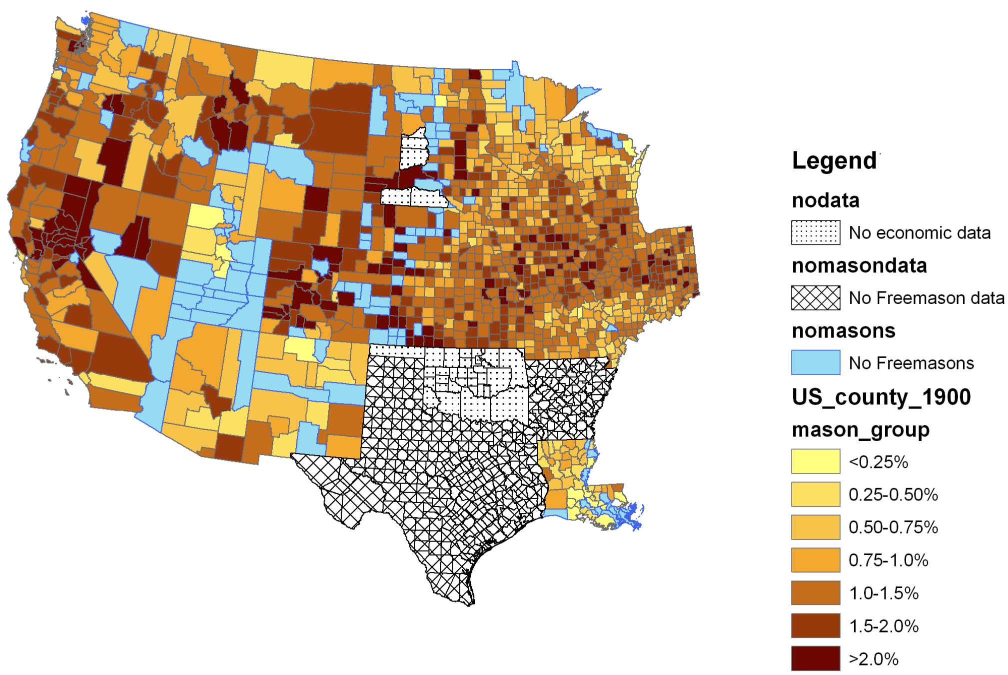


Table 2.1: Freemasonry and Education in the Late-19th Century

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | **1870** | | | | | **1890** | | |
| | | | Share of Output from | | | | | Share of Output from | | |
| | | | < 33 % | 33-66% | > 66% | | | < 33 % | 33-66% | > 66% |
| log($Y^M$ per capita) | 0.03** | 0.02** | 0.03* | 0.00 | 0.01 | -0.01 | -0.01 | -0.02 | -0.07*** | 0.05 |
| | (0.01) | (0.01) | (0.02) | (0.03) | (0.03) | (0.01) | (0.01) | (0.01) | (0.02) | (0.03) |
| log($Y^A$ per capita) | 0.00 | 0.01 | 0.02 | -0.04 | 0.00 | 0.01 | 0.02 | 0.01 | 0.04 | 0.04 |
| | (0.01) | (0.01) | (0.02) | (0.04) | (0.03) | (0.01) | (0.01) | (0.01) | (0.03) | (0.03) |
| log($K^M$ per capita) | 0.01 | 0.00 | 0.00 | -0.00 | 0.07** | 0.00 | 0.00 | 0.00 | 0.04** | -0.03 |
| | (0.01) | (0.01) | (0.02) | (0.02) | (0.03) | (0.01) | (0.01) | (0.01) | (0.02) | (0.03) |
| log($K^A$ per capita) | 0.02** | 0.03** | 0.03** | 0.07*** | -0.05 | 0.01 | 0.00 | 0.02 | 0.00 | -0.02 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.04) | (0.01) | (0.01) | (0.02) | (0.03) | (0.04) |
| Proportion of population native born | 0.17** | 0.13* | 0.00 | 0.18* | 0.38** | 0.43*** | 0.39*** | 0.42*** | 0.48*** | 0.20 |
| | (0.07) | (0.07) | (0.08) | (0.11) | (0.19) | (0.05) | (0.05) | (0.06) | (0.09) | (0.13) |
| Proportion of population in towns with < 25,000 people | -0.00 | 0.04 | 0.00 | 0.00 | 0.02 | -0.08*** | -0.07** | 0.00 | 0.00 | -0.07* |
| | (0.09) | (0.08) | . | . | (0.10) | (0.03) | (0.03) | . | . | (0.04) |
| log(population) | 0.03*** | 0.02*** | 0.02* | 0.02 | 0.02 | -0.01 | -0.00 | 0.01 | -0.02 | -0.01 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| % of Men over 21 | -0.07 | -0.07 | -0.47** | 0.13 | 0.29 | -0.37*** | -0.41*** | -0.65*** | -0.38 | 0.33 |
| | (0.12) | (0.12) | (0.20) | (0.20) | (0.23) | (0.13) | (0.12) | (0.12) | (0.24) | (0.23) |
| # Masons / Total Population | | 3.77*** | 4.01*** | 3.22** | 2.03 | | 3.63*** | 3.54*** | 3.78*** | 3.17 |
| | | (0.75) | (0.93) | (1.48) | (1.76) | | (0.74) | (0.90) | (1.25) | (2.39) |
| $R^2$ | 0.59 | 0.61 | 0.62 | 0.66 | 0.44 | 0.40 | 0.41 | 0.41 | 0.53 | 0.40 |
| N = | 1012 | 1012 | 660 | 231 | 121 | 1304 | 1304 | 849 | 285 | 170 |

Notes: Columns (3) and (8) include counties with $\leq$ 33% of output from manufacturing, (4) and (9) are counties with 34-66% of output from manufacturing and (5) and (10) include counties with $\geq$ 67% of output from manufacturing.
All specifications include state fixed effects. Robust standard errors in parentheses.
1: For 1890, this variable is the "Population of Population Native Born, with Native Parentage"

# Table 2.2: Lasting Impact of Freemasonry on American Education

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Enrollment Among Youth of the Indicated Age Group | | | | | | | | | | | |
| | 7-13 | | | 14-15 | | | 16-17 | | | 18-20 | | |
| Proportion of population in towns: | | | | | | | | | | | | |
| with < 2,500 people | -0.01*** | -0.01** | -0.01** | -0.07*** | -0.07*** | -0.07*** | -0.11*** | -0.12*** | -0.12*** | -0.07*** | -0.08*** | -0.08*** |
| | (0.00) | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| with 2,500-25,000 people | -0.00 | -0.00 | -0.00 | -0.03*** | -0.03*** | -0.04*** | -0.05*** | -0.05*** | -0.05*** | -0.01 | -0.02* | -0.02** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Proportion of population native born [1940] | -0.06** | -0.06** | -0.06** | -0.09*** | -0.08*** | -0.09*** | -0.04 | -0.04 | -0.05 | 0.03 | 0.02 | 0.02 |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) | (0.02) | (0.02) | (0.02) |
| log(Value of crops per farm) [1940] | 0.01** | 0.01** | 0.01** | 0.00 | 0.01 | 0.01* | 0.01 | 0.01 | 0.01** | 0.00 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) | (0.00) | (0.00) | (0.00) |
| log($Y^M$ per capita) | | 0.01*** | 0.01** | | 0.01*** | 0.01** | | 0.01** | 0.01 | | 0.00 | 0.00 |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.01) | | (0.00) | (0.00) |
| log($Y^A$ per capita) | | -0.00 | -0.00 | | -0.00 | -0.00 | | -0.00 | -0.00 | | -0.00 | -0.00 |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.00) |
| log($K^M$ per capita) | | -0.01** | -0.01** | | -0.01** | -0.01** | | -0.01* | -0.01** | | -0.01* | -0.01** |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.01) | (0.01) | | (0.00) | (0.00) |
| log($K^A$ per capita) | | 0.00 | 0.00 | | 0.00 | 0.00 | | 0.00 | 0.00 | | 0.00 | 0.00 |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.01) | (0.00) | | (0.00) | (0.00) |
| Proportion of population in towns with < 25,000 people [1870] | | 0.00 | 0.01 | | 0.02 | 0.03 | | 0.04 | 0.05 | | -0.00 | 0.00 |
| | | (0.01) | (0.01) | | (0.02) | (0.02) | | (0.03) | (0.03) | | (0.02) | (0.02) |
| Proportion of population native born [1870] | | 0.05*** | 0.05*** | | 0.13*** | 0.13*** | | 0.18*** | 0.17*** | | 0.09*** | 0.09*** |
| | | (0.01) | (0.01) | | (0.02) | (0.02) | | (0.03) | (0.03) | | (0.01) | (0.01) |
| log(population) [1870] | | 0.00 | 0.00 | | 0.00* | 0.00 | | 0.00 | 0.00 | | 0.00 | 0.00 |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.00) |
| % of Men over 21 [1870] | | 0.07*** | 0.07*** | | 0.14*** | 0.14*** | | 0.22*** | 0.23*** | | 0.06** | 0.06*** |
| | | (0.02) | (0.02) | | (0.03) | (0.03) | | (0.04) | (0.04) | | (0.02) | (0.02) |
| # Masons / Total Population [1870] | | | 0.13 | | | 0.92*** | | | 1.94*** | | | 0.69*** |
| | | | (0.17) | | | (0.25) | | | (0.38) | | | (0.21) |
| $R^2$ | 0.30 | 0.32 | 0.32 | 0.42 | 0.46 | 0.47 | 0.49 | 0.53 | 0.55 | 0.45 | 0.47 | 0.48 |
| N = | 1030 | 1014 | 1014 | 1030 | 1014 | 1014 | 1030 | 1014 | 1014 | 1030 | 1014 | 1014 |

Note: All specifications include state fixed effects. Robust standard errors in parentheses.

Table 2.3: Panel Data Estimation using 1850-1900 Data (T = 6, 11 states included)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Type of Estimator: | OLS | FE | AH (1982) | | AB (1991) |
| Type of IV: | | | Level | Difference | GMM |
| $\log(Y^M$ per capita) | 0.01 | 0.01 | 0.00 | -0.00 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| $\log(K^M$ per capita) | -0.01 | -0.01 | -0.00 | -0.00 | -0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| $\log(K^A$ per capita) | 0.04*** | 0.03*** | 0.03*** | 0.04*** | 0.03*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Proportion of population in towns with < | -0.05 | -0.11*** | -0.03 | -0.01 | -0.06 |
| 25,000 people | (0.03) | (0.03) | (0.05) | (0.05) | (0.05) |
| log(population) | 0.03*** | 0.08*** | 0.10*** | 0.11*** | 0.10*** |
| | (0.00) | (0.01) | (0.01) | (0.02) | (0.01) |
| % of Men over 21 | -0.25** | -0.54*** | -0.78*** | -1.35*** | -0.76*** |
| | (0.10) | (0.11) | (0.09) | (0.16) | (0.21) |
| # Masons / Total Population | 2.65*** | 0.90 | 2.20** | 1.82 | 1.89** |
| | (0.74) | (0.72) | (0.94) | (1.70) | (0.90) |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes |
| State Fixed Effects | Yes | NA | NA | NA | NA |
| $R^2$ | 0.55 | 0.42 | 0.40 | 0.43 | |
| N = | 2552 | 2552 | 2011 | 1561 | 2011 |

Table 2.4: Panel Data Estimation using 1860-1900 Data (T = 5, 16 states included)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Type of Estimator: | OLS | FE | AH (1982) | | AB (1991) |
| Type of IV: | | | Level | Difference | GMM |
| $\log(Y^M$ per capita) | 0.01 | -0.00 | 0.01 | 0.01* | 0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| $\log(K^M$ per capita) | -0.00 | 0.00 | -0.01 | -0.02** | -0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| $\log(K^A$ per capita) | 0.02*** | 0.01** | 0.02*** | 0.01** | 0.02** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Proportion of population native born | -0.06** | -0.12*** | -0.06 | -0.07 | -0.07** |
| | (0.03) | (0.03) | (0.05) | (0.05) | (0.04) |
| Proportion of population in towns with < 25,000 people | 0.04 | -0.07 | -0.32*** | -0.16 | -0.30*** |
| | (0.03) | (0.05) | (0.04) | (0.12) | (0.07) |
| log(population) | 0.02*** | 0.04*** | 0.03*** | 0.05*** | 0.04*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| % of Men over 21 | -0.49*** | -0.75*** | -0.71*** | -0.87*** | -0.75*** |
| | (0.09) | (0.13) | (0.12) | (0.19) | (0.18) |
| # Masons / Total Population | 3.61*** | 1.71** | 3.15*** | 1.66 | 2.07** |
| | (0.51) | (0.75) | (1.06) | (1.59) | (1.01) |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes |
| State Fixed Effects | Yes | NA | NA | NA | NA |
| $R^2$ | 0.46 | 0.23 | 0.30 | 0.34 | |
| N = | 3818 | 3818 | 2962 | 2198 | 2962 |

Table 2.5: Panel Data Estimation using 1870-1900 Data (T = 4, 23 states included)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Type of Estimator: | OLS | FE | AH (1982) | | AB (1991) |
| Type of IV: | | | Level | Difference | GMM |
| log($Y^M$ per capita) | 0.01 | 0.01 | 0.01** | 0.02 | 0.02 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($Y^A$ per capita) | -0.00 | -0.01 | -0.00 | 0.03*** | -0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($K^M$ per capita) | -0.01 | -0.01 | -0.02** | -0.02** | -0.02 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($K^A$ per capita) | 0.02*** | 0.01* | 0.01 | 0.02 | 0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Proportion of population native born | 0.13*** | 0.04 | 0.12 | -0.45*** | 0.13 |
| | (0.04) | (0.09) | (0.09) | (0.15) | (0.13) |
| Proportion of population in towns with < 25,000 people | -0.10*** | -0.14*** | -0.11** | -0.02 | -0.13*** |
| | (0.02) | (0.03) | (0.05) | (0.05) | (0.04) |
| log(population) | 0.03*** | 0.06*** | 0.05*** | 0.01 | 0.06*** |
| | (0.00) | (0.01) | (0.01) | (0.02) | (0.01) |
| % of Men over 21 | -0.30*** | -0.52*** | -0.46*** | -1.21*** | -0.49*** |
| | (0.08) | (0.12) | (0.10) | (0.18) | (0.16) |
| # Masons / Total Population | 3.02*** | -0.07 | 0.91 | -2.48 | -0.34 |
| | (0.48) | (0.63) | (0.92) | (2.43) | (0.86) |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes |
| State Fixed Effects | Yes | NA | NA | NA | NA |
| $R^2$ | 0.50 | 0.29 | 0.30 | 0.38 | |
| N = | 3885 | 3885 | 2819 | 1821 | 2819 |

Table 2.6: Freemasonry and Social Heterogeneity in 1870

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Share of Output from Manufacturing | | | | | |
| | | | | < 33 % | | 34-66% | | > 66% | |
| log($Y^M$ per capita) | 0.03** | 0.02** | 0.02** | 0.03* | 0.03* | 0.00 | 0.00 | 0.01 | 0.01 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.03) | (0.03) | (0.03) | (0.03) |
| log($Y^A$ per capita) | 0.00 | 0.01 | 0.00 | 0.02 | 0.02 | -0.04 | -0.04 | 0.00 | 0.02 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.04) | (0.04) | (0.03) | (0.03) |
| log($K^M$ per capita) | 0.01 | 0.00 | 0.00 | 0.00 | -0.00 | -0.00 | 0.00 | 0.07** | 0.06** |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.02) | (0.02) | (0.03) | (0.03) |
| log($K^A$ per capita) | 0.02** | 0.03** | 0.03*** | 0.03** | 0.03** | 0.07*** | 0.07*** | -0.05 | -0.04 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.04) | (0.04) |
| Proportion of population native born | 0.17** | 0.13* | 0.20** | 0.00 | 0.10 | 0.18* | 0.33** | 0.38** | 0.46** |
| | (0.07) | (0.07) | (0.08) | (0.08) | (0.10) | (0.11) | (0.15) | (0.19) | (0.20) |
| Proportion of population in towns with < 25,000 people | -0.00 | 0.04 | 0.04 | | | | | 0.02 | 0.04 |
| | (0.09) | (0.08) | (0.08) | | | | | (0.10) | (0.10) |
| log(population) | 0.03*** | 0.02*** | 0.02*** | 0.02* | 0.02* | 0.02 | 0.01 | 0.02 | 0.02 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.02) | (0.02) |
| % of Men over 21 | -0.07 | -0.07 | -0.08 | -0.47** | -0.50** | 0.13 | 0.07 | 0.29 | 0.29 |
| | (0.12) | (0.12) | (0.12) | (0.20) | (0.20) | (0.20) | (0.19) | (0.23) | (0.24) |
| # Masons / Total Population | | 3.77*** | 10.47*** | 4.01*** | 19.96*** | 3.22** | 14.54** | 2.03 | 14.47** |
| | | (0.75) | (2.72) | (0.93) | (6.31) | (1.48) | (7.01) | (1.76) | (6.39) |
| (% Masons) X (% Native) | | | -8.62** | | -18.28** | | -14.69 | | -19.67** |
| | | | (3.49) | | (7.12) | | (9.01) | | (9.70) |
| $R^2$ | 0.59 | 0.61 | 0.61 | 0.62 | 0.62 | 0.66 | 0.66 | 0.44 | 0.46 |
| N = | 1012 | 1012 | 1012 | 660 | 660 | 231 | 231 | 121 | 121 |

Table 2.7: Ethnic Homogeneity and Freemasonry - Panel Data Estimates

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1860-1900 Panel Data | | | | | | 1870-1900 Panel Data | | | | | |
| | OLS | | | FE | | | Pooled OLS | | | FE | | |
| log($Y^M$ per capita) | 0.01 | 0.01 | 0.01 | 0.00 | -0.00 | -0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($Y^A$ per capita) | | | | | | | -0.00 | -0.00 | -0.00 | -0.01 | -0.01 | -0.01 |
| | | | | | | | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($K^M$ per capita) | -0.00 | -0.00 | -0.00 | 0.00 | 0.00 | 0.00 | -0.01 | -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($K^A$ per capita) | 0.02*** | 0.02*** | 0.02*** | 0.02** | 0.01** | 0.01* | 0.02*** | 0.02*** | 0.02*** | 0.01* | 0.01* | 0.01** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Proportion of population native born | 0.06* | 0.04 | 0.09** | -0.07 | -0.06 | 0.03 | 0.16*** | 0.12*** | 0.20*** | 0.04 | 0.04 | 0.13 |
| | (0.03) | (0.03) | (0.04) | (0.05) | (0.05) | (0.06) | (0.04) | (0.04) | (0.04) | (0.09) | (0.09) | (0.09) |
| Proportion of population in towns with < 25,000 people | -0.06** | -0.06** | -0.06** | -0.13*** | -0.12*** | -0.11*** | -0.11*** | -0.10*** | -0.10*** | -0.14*** | -0.14*** | -0.13*** |
| | (0.03) | (0.03) | (0.03) | (0.04) | (0.03) | (0.03) | (0.02) | (0.02) | (0.02) | (0.03) | (0.03) | (0.03) |
| log(population) | 0.02*** | 0.02*** | 0.02*** | 0.04*** | 0.04*** | 0.04*** | 0.03*** | 0.03*** | 0.02*** | 0.06*** | 0.06*** | 0.05*** |
| | (0.00) | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) | (0.00) | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) |
| % of Men over 21 | -0.45*** | -0.48*** | -0.49*** | -0.75*** | -0.73*** | -0.75*** | -0.28*** | -0.30*** | -0.31*** | -0.52*** | -0.52*** | -0.59*** |
| | (0.09) | (0.09) | (0.09) | (0.13) | (0.13) | (0.13) | (0.08) | (0.08) | (0.08) | (0.12) | (0.12) | (0.12) |
| # Masons / Total Population | | 3.98*** | 9.65*** | | 2.05** | 14.36*** | | 3.51*** | 11.74*** | | 0.02 | 17.92*** |
| | | (0.53) | (2.97) | | (0.92) | (4.30) | | (0.51) | (2.20) | | (0.78) | (3.92) |
| (% Masons) X (% Native) | | | -6.78** | | | -14.60*** | | | -10.18*** | | | -21.87*** |
| | | | (3.39) | | | (4.94) | | | (2.56) | | | (4.61) |
| $R^2$ | 0.45 | 0.46 | 0.46 | 0.23 | 0.23 | 0.24 | 0.49 | 0.50 | 0.50 | 0.29 | 0.29 | 0.30 |
| N = | 3818 | 3818 | 3818 | 3818 | 3818 | 3818 | 3885 | 3885 | 3885 | 3885 | 3885 | 3885 |

Table 2.8: Ethnic Fractionalization - Panel Data Estimates

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pooled OLS | | | | | | | Fixed Effects | | | | | | |
| log($Y^M$ per capita) | 0.02* | 0.02** | 0.02* | 0.02** | 0.02** | 0.02* | 0.02** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($Y^A$ per capita) | 0.00 | 0.00 | -0.00 | -0.00 | 0.00 | 0.00 | 0.00 | -0.01* | -0.01 | -0.01 | -0.01* | -0.01* | -0.01* | -0.01* |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($K^M$ per capita) | -0.01 | -0.02 | -0.02* | -0.02* | -0.02* | -0.02* | -0.02* | -0.01 | -0.01 | -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| log($K^A$ per capita) | 0.01*** | 0.02*** | 0.02*** | 0.02*** | 0.01*** | 0.02*** | 0.02*** | 0.02*** | 0.02*** | 0.02*** | 0.02*** | 0.02*** | 0.02*** | 0.02*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Proportion of population in towns with < 25.000 people | -0.13*** | -0.12*** | -0.10*** | -0.10*** | -0.13*** | -0.11*** | -0.10*** | -0.16*** | -0.15*** | -0.15*** | -0.13*** | -0.15*** | -0.15*** | -0.13*** |
| | (0.03) | (0.03) | (0.02) | (0.02) | (0.03) | (0.02) | (0.02) | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) |
| log(population) | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.06*** | 0.07*** | 0.07*** | 0.06*** | 0.06*** | 0.06*** | 0.06*** |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| % of Men over 21 | -0.23*** | -0.28*** | -0.31*** | -0.33*** | -0.31*** | -0.32*** | -0.32*** | -0.51*** | -0.53*** | -0.53*** | -0.61*** | -0.52*** | -0.52*** | -0.54*** |
| | (0.08) | (0.09) | (0.09) | (0.09) | (0.08) | (0.08) | (0.08) | (0.11) | (0.11) | (0.11) | (0.12) | (0.10) | (0.10) | (0.10) |
| % Native Born | 0.17*** | | | | | | | 0.02 | | | | | | |
| | (0.04) | | | | | | | (0.09) | | | | | | |
| Proportion of population native born, native parentage | | 0.09* | 0.05 | 0.12** | | | | | 0.04 | 0.04 | 0.14 | | | |
| | | (0.05) | (0.05) | (0.06) | | | | | (0.09) | (0.09) | (0.09) | | | |
| Proportion of population native born, at least one foreign parent | | -0.04 | -0.06 | -0.05 | | | | | -0.18 | -0.18 | -0.11 | | | |
| | | (0.09) | (0.09) | (0.10) | | | | | (0.11) | (0.11) | (0.12) | | | |
| # Masons / Total Population | | | 3.86*** | 11.27*** | | 4.06*** | 1.61* | | | 0.28 | 18.66*** | | 0.25 | -6.86*** |
| | | | (0.51) | (2.62) | | (0.50) | (0.96) | | | (0.81) | (3.86) | | (0.82) | (1.72) |
| (% Masons) X (% Native Parentage) | | | | -9.87*** | | | | | | | -23.45*** | | | |
| | | | | (2.73) | | | | | | | (4.52) | | | |
| (% Masons) X (% Foreign Parentage) | | | | -6.23 | | | | | | | -17.72** | | | |
| | | | | (5.48) | | | | | | | (8.02) | | | |
| EFI | | | | | -0.04** | -0.03 | -0.09*** | | | | | -0.01 | -0.01 | -0.12 |
| | | | | | (0.02) | (0.02) | (0.03) | | | | | (0.07) | (0.07) | (0.07) |
| (% Masons) X (EFI) | | | | | | | 5.32*** | | | | | | | 16.19*** |
| | | | | | | | (1.74) | | | | | | | (3.21) |
| $R^2$ | 0.47 | 0.48 | 0.49 | 0.49 | 0.47 | 0.49 | 0.49 | 0.32 | 0.33 | 0.33 | 0.34 | 0.32 | 0.32 | 0.34 |
| N = | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 | 2931 |

*: Significance at 10% level, **: Significant at 5% level, ***: Significance at the 1% level.

Table 2.9: Religious Homogeneity and Freemasonry - Panel Data Estimates

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pooled OLS | | | | | Fixed Effects | | | | | AB (1991) | | | | |
| log($Y^M$ per capita) | 0.01 | 0.01 | 0.00 | -0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | -0.00 | 0.00 | -0.00 | -0.00 | -0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) |
| log($K^M$ per capita) | -0.00 | -0.01 | -0.00 | 0.00 | -0.00 | -0.00 | -0.02 | -0.00 | -0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) |
| log($K^A$ per capita) | 0.03*** | 0.02*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.03*** | 0.02*** | 0.03*** | 0.03** | 0.03** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Proportion of population in towns with < 25,000 people | -0.08** | -0.10** | -0.07 | -0.04 | -0.04 | -0.21*** | -0.14** | -0.15*** | -0.13** | -0.12** | -0.16*** | 0.12 | -0.11 | -0.10 | -0.12* |
| | (0.04) | (0.05) | (0.04) | (0.05) | (0.05) | (0.04) | (0.06) | (0.05) | (0.05) | (0.05) | (0.05) | (0.16) | (0.06) | (0.06) | (0.07) |
| log(population) | 0.03*** | 0.03*** | 0.03*** | 0.01** | 0.01** | 0.09*** | 0.07*** | 0.09*** | 0.07*** | 0.07*** | 0.11*** | 0.09*** | 0.11*** | 0.10*** | 0.10*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.02) | (0.02) |
| % of Men over 21 | -0.22** | -0.28*** | -0.21* | -0.18 | -0.19 | -0.55*** | -0.47*** | -0.56*** | -0.65*** | -0.65*** | -0.75*** | -0.51*** | -0.76*** | -0.83*** | -0.79*** |
| | (0.11) | (0.11) | (0.11) | (0.14) | (0.14) | (0.13) | (0.13) | (0.13) | (0.17) | (0.17) | (0.22) | (0.18) | (0.22) | (0.24) | (0.24) |
| # Masons per capita | 2.50*** | 2.20*** | 2.39*** | 2.25*** | 1.17 | -0.03 | 0.27 | -0.32 | -0.32 | 0.95 | 2.52** | 1.63 | 2.25* | 1.91 | 0.59 |
| | (0.89) | (0.82) | (0.88) | (0.85) | (0.77) | (0.97) | (1.03) | (0.97) | (0.98) | (1.05) | (1.21) | (1.31) | (1.21) | (1.33) | (1.11) |
| # Catholic seats per capita | | -0.17*** | | | | | -0.09 | | | | | 0.06 | | | |
| | | (0.06) | | | | | (0.06) | | | | | (0.04) | | | |
| # Jewish seats per capita | | -3.09 | -1.44 | -1.57 | -1.60 | | -6.63* | -6.05* | -6.16* | -5.69* | | -12.56 | -5.11 | -5.40 | -6.27* |
| | | (2.92) | (2.84) | (2.87) | (2.87) | | (3.61) | (3.25) | (3.24) | (3.20) | | (8.17) | (3.89) | (3.84) | (3.79) |
| # seats in all churches per capita | | 0.04** | 0.02 | 0.01 | 0.01 | | 0.04** | 0.03** | 0.03* | 0.03* | | -0.01 | 0.04* | 0.04* | 0.04* |
| | | (0.02) | (0.01) | (0.01) | (0.01) | | (0.02) | (0.01) | (0.02) | (0.02) | | (0.03) | (0.02) | (0.02) | (0.02) |
| RFI | | | | 0.10*** | 0.08*** | | | | 0.03 | 0.05 | | | | 0.02 | -0.03 |
| | | | | (0.03) | (0.03) | | | | (0.03) | (0.04) | | | | (0.05) | (0.06) |
| (RFI) X (# Masons per capita) | | | | | 2.54* | | | | | -3.30* | | | | | 8.11* |
| | | | | | (1.49) | | | | | (2.00) | | | | | (4.48) |
| $R^2$ | 0.56 | 0.57 | 0.56 | 0.55 | 0.55 | 0.49 | 0.55 | 0.49 | 0.38 | 0.39 | | | | | |
| N = | 1728 | 1273 | 1728 | 1650 | 1650 | 1728 | 1273 | 1728 | 1650 | 1650 | 1183 | 363 | 1183 | 1119 | 1119 |

*: Significance at 10% level, **: Significant at 5% level, ***: Significance at the 1% level.
AB (1991) instruments for both Freemasons per capita and the interact of Freemasons per capita with RFI.

Table 2.10: Freemasonry and Educational Expenditures in the Mid-19th Century (1850)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | # Teachers per Potential Student[1] | | | | Public Education Income from | | | |
| | Public | | Private (Academy) | | Taxation | | Public Funds | |
| $\log(Y^M$ per capita) | 0.00** | 0.00** | -0.00** | -0.00** | 0.00 | 0.00 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.02) | (0.02) | (0.01) | (0.01) |
| $\log(K^M$ per capita) | -0.00** | -0.00** | 0.00** | 0.00** | -0.00 | -0.00 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.02) | (0.02) | (0.01) | (0.01) |
| $\log(K^A$ per capita) | 0.00*** | 0.00*** | 0.00 | -0.00 | 0.01*** | 0.01*** | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Proportion of population in towns with < 2,500 people | 0.00 | 0.00 | -0.00*** | -0.00*** | -0.13 | -0.14* | 0.02 | 0.02 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.08) | (0.08) | (0.09) | (0.09) |
| Proportion of population in towns with 2,500-25,000 people | -0.00 | -0.00 | 0.00 | 0.00 | 0.15 | 0.14 | 0.01 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.23) | (0.23) | (0.10) | (0.10) |
| log(population) | 0.00* | 0.00* | -0.00 | -0.00 | 0.01** | 0.01** | 0.01 | 0.01 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) | (0.01) |
| % of Men over 21 | -0.00 | -0.00 | 0.00 | -0.00 | 0.02 | 0.01 | -0.06 | -0.07 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.10) | (0.10) | (0.06) | (0.06) |
| # Masons / Total Population | | -0.05 | | 0.13*** | | 1.73 | | 2.01 |
| | | (0.07) | | (0.05) | | (3.06) | | (2.17) |
| $R^2$ | 0.36 | 0.36 | 0.42 | 0.45 | 0.22 | 0.22 | 0.06 | 0.06 |
| N = | 482 | 482 | 482 | 482 | 482 | 482 | 482 | 482 |

Notes:
[1]: A potential student is everyone between the ages of 5-19. Changing this to the total population only affects the size of the point estimates.

Table 2.11: Freemasonry and Taxation in 1870 by Locality

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Local Taxes ($/Person) | | | | | County Taxes ($/Person) | | | | | State Taxes ($/Person) | | | | |
| | | | Share of Output from Manufacturing | | | | | Share of Output from Manufacturing | | | | | Share of Output from Manufacturing | | |
| | | | < 33 % | 33-66% | > 66% | | | < 33 % | 33-66% | > 66% | | | < 33 % | 33-66% | > 66% |
| log($Y^M$ per capita) | -0.05 | -0.11 | 0.02 | -0.67 | -0.05 | 0.27 | 0.20 | 0.51 | -0.23 | 0.74 | -0.05 | -0.09 | 0.08 | -0.43 | -0.48 |
| | (0.11) | (0.12) | (0.13) | (0.88) | (0.62) | (0.18) | (0.18) | (0.27) | (0.55) | (0.77) | (0.09) | (0.09) | (0.11) | (0.27) | (0.40) |
| log($Y^A$ per capita) | 0.02 | -0.00 | 0.04 | 0.30 | 0.32 | 0.21 | 0.18 | 0.07 | 0.24 | -0.21 | 0.13 | 0.11 | 0.29 | 0.23 | 0.27 |
| | (0.09) | (0.09) | (0.11) | (0.81) | (0.45) | (0.16) | (0.16) | (0.22) | (0.54) | (0.77) | (0.09) | (0.09) | (0.19) | (0.29) | (0.40) |
| log($K^M$ per capita) | 0.32* | 0.31* | 0.17 | 0.51 | 0.01 | 0.11 | 0.11 | -0.21 | 0.47 | 0.12 | 0.15 | 0.14 | -0.09 | 0.25* | 0.40 |
| | (0.14) | (0.13) | (0.12) | (0.50) | (0.62) | (0.14) | (0.14) | (0.19) | (0.31) | (0.51) | (0.08) | (0.08) | (0.09) | (0.12) | (0.29) |
| log($K^A$ per capita) | 0.07 | 0.13 | 0.38** | -0.00 | -0.75 | 0.10 | 0.16 | 0.16 | 0.39 | 0.04 | 0.26* | 0.29* | 0.36* | 0.38* | -0.19 |
| | (0.09) | (0.11) | (0.14) | (0.33) | (0.47) | (0.19) | (0.19) | (0.25) | (0.34) | (0.73) | (0.12) | (0.12) | (0.17) | (0.18) | (0.43) |
| % Native | -0.21 | -0.55 | -0.23 | 1.49 | 3.02 | -0.28 | -0.63 | -1.43 | 2.79 | -0.44 | -0.17 | -0.34 | -0.93 | 0.68 | -0.89 |
| | (0.55) | (0.50) | (0.70) | (1.22) | (4.26) | (0.80) | (0.82) | (1.11) | (1.48) | (4.60) | (0.53) | (0.56) | (0.72) | (0.59) | (2.59) |
| % in Towns < 25,000 people | 5.84** | 6.13** | 0.00 | 0.00 | 2.81 | 4.99* | 5.30* | 0.00 | 0.00 | 4.17 | 3.03** | 3.19** | 0.00 | 0.00 | 2.88* |
| | (2.26) | (2.33) | . | . | (3.24) | (2.46) | (2.40) | . | . | (2.46) | (0.84) | (0.84) | . | . | (1.17) |
| log(population) | -0.13 | -0.16 | -0.43 | 0.04 | 0.24 | -0.45 | -0.50 | -0.46 | -0.21 | -0.40 | -0.25* | -0.28* | -0.25 | -0.05 | -0.33 |
| | (0.15) | (0.16) | (0.32) | (0.22) | (0.40) | (0.27) | (0.27) | (0.43) | (0.26) | (0.47) | (0.13) | (0.13) | (0.19) | (0.09) | (0.24) |
| % of Men over 21 | 1.17 | 1.24 | 2.92 | 1.56 | 6.49 | 3.51* | 3.60* | 4.90 | 8.42** | -3.25 | 0.98 | 1.02 | 2.14 | 2.12* | -1.21 |
| | (0.72) | (0.73) | (1.50) | (1.25) | (5.21) | (1.77) | (1.74) | (3.66) | (2.55) | (5.57) | (0.92) | (0.90) | (2.06) | (1.06) | (3.23) |
| # Masons / Total Population | | 34.99 | 59.66 | 6.67 | 34.25 | | 36.83** | 17.16 | 4.42 | 72.58** | | 18.49** | 8.77 | 7.27 | 43.39* |
| | | (21.86) | (50.18) | (11.93) | (23.78) | | (11.14) | (17.70) | (20.50) | (22.37) | | (6.86) | (7.02) | (10.89) | (20.50) |
| $R^2$ | 0.32 | 0.33 | 0.38 | 0.26 | 0.31 | 0.36 | 0.37 | 0.30 | 0.51 | 0.65 | 0.39 | 0.40 | 0.38 | 0.59 | 0.65 |
| N = | 1173 | 1173 | 747 | 297 | 129 | 1247 | 1247 | 803 | 298 | 146 | 1247 | 1247 | 803 | 298 | 146 |

Notes:
Columns (3) and (8) include counties with $\leq$ 33% of output from manufacturing, (4) and (9) are counties with 34-66% of output from manufacturing and (5) and (10) include counties with $\geq$ 67% of output from manufacturing.
All specifications include state fixed effects. Robust standard errors in parentheses.

Table 2.12: Freemasonry and the Share of Local to State Taxes

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Local as a Share of Total Taxes** | | | | | **County as a Share of Total Taxes** | | | | |
| | | | Share of Output from Manufacturing | | | | | Share of Output from Manufacturing | | |
| | | | < 33% | 33-66% | > 66% | | | < 33% | 33-66% | > 66% |
| $\log(Y^M$ per capita) | 0.00 | 0.00 | 0.00 | -0.03 | -0.00 | 0.02* | 0.02* | 0.01 | 0.05 | 0.04 |
| | (0.01) | (0.01) | (0.01) | (0.04) | (0.04) | (0.01) | (0.01) | (0.01) | (0.03) | (0.03) |
| $\log(Y^A$ per capita) | -0.00 | -0.00 | 0.00 | 0.02 | 0.02 | 0.01 | 0.01 | -0.02 | -0.03 | -0.01 |
| | (0.01) | (0.01) | (0.01) | (0.04) | (0.02) | (0.01) | (0.01) | (0.02) | (0.03) | (0.02) |
| $\log(K^M$ per capita) | 0.01 | 0.01 | -0.00 | 0.01 | 0.02 | -0.02** | -0.02* | -0.01 | -0.02 | -0.02 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.04) | (0.01) | (0.01) | (0.01) | (0.02) | (0.03) |
| $\log(K^A$ per capita) | 0.01 | 0.01 | 0.02* | -0.00 | 0.01 | -0.03*** | -0.03*** | -0.03** | -0.03 | -0.02 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.03) | (0.01) | (0.01) | (0.01) | (0.02) | (0.03) |
| % Native | -0.09 | -0.10* | -0.13 | -0.06 | -0.19 | 0.02 | 0.03 | 0.01 | -0.07 | 0.65*** |
| | (0.05) | (0.05) | (0.08) | (0.10) | (0.13) | (0.06) | (0.06) | (0.08) | (0.09) | (0.17) |
| % in Towns < 25,000 people | 0.08 | 0.09 | | | 0.10 | -0.08 | -0.08 | | | -0.14* |
| | (0.10) | (0.10) | | | (0.09) | (0.11) | (0.11) | | | (0.08) |
| log(population) | 0.03*** | 0.03*** | 0.03*** | 0.02 | 0.01 | -0.01 | -0.01 | -0.01 | -0.01 | 0.05*** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| % of Men over 21 | 0.00 | 0.01 | 0.05 | -0.09 | -0.00 | 0.00 | -0.00 | -0.15 | 0.10 | 0.54** |
| | (0.06) | (0.06) | (0.15) | (0.11) | (0.13) | (0.11) | (0.11) | (0.19) | (0.11) | (0.22) |
| # Masons / Total Population | | 1.12** | 1.33 | 0.47 | 0.49 | | -0.73 | 0.10 | -1.35 | -0.50 |
| | | (0.54) | (0.85) | (0.96) | (0.84) | | (0.58) | (0.85) | (0.98) | (1.00) |
| $R^2$ | 0.69 | 0.69 | 0.71 | 0.69 | 0.66 | 0.46 | 0.46 | 0.42 | 0.60 | 0.63 |
| N = | 939 | 939 | 616 | 228 | 95 | 939 | 939 | 616 | 228 | 95 |

Table 2.13: Correlates of Freemasonry during the 19th Century

| | (1) | (2) | (3) | (4) | (5) | (6) | (8) | (9) | (10) | (11) |
|---|---|---|---|---|---|---|---|---|---|---|
| | **1850** | **1860** | **1870** | **1880** | **1890** | **1900** | **1850** | **1860** | **1870** | **1890** |
| log($Y^M$ per capita) | -0.05 | -0.01 | 0.00 | -0.05 | -0.01 | 0.09 | -0.02 | 0.03 | -0.04 | 0.01 |
| | (0.05) | (0.04) | (0.06) | (0.05) | (0.05) | (0.06) | (0.03) | (0.04) | (0.06) | (0.05) |
| log($Y^A$ per capita) | | | -0.03 | 0.07** | -0.07 | 0.38*** | 0.02 | -0.02 | 0.08 | 0.00 |
| | | | (0.03) | (0.03) | (0.12) | (0.06) | (0.03) | (0.05) | (0.07) | (0.05) |
| log($K^M$ per capita) | 0.05 | 0.03 | 0.06 | 0.10* | 0.02 | 0.02 | | | 0.01 | -0.12 |
| | (0.05) | (0.04) | (0.07) | (0.05) | (0.05) | (0.06) | | | (0.06) | (0.11) |
| log($K^A$ per capita) | 0.04* | 0.04** | -0.02 | -0.06* | 0.03 | -0.33*** | 0.07*** | 0.04 | -0.07 | 0.06 |
| | (0.02) | (0.02) | (0.03) | (0.03) | (0.13) | (0.06) | (0.03) | (0.06) | (0.05) | (0.13) |
| Proportion of population native | | 0.27 | 0.66*** | 0.67*** | 1.12*** | 1.30*** | | 0.02 | 0.80** | 0.98*** |
| born | | (0.17) | (0.25) | (0.24) | (0.24) | (0.31) | | (0.26) | (0.36) | (0.26) |
| Proportion of population in towns | -0.23** | -0.05 | -0.63*** | -0.32** | -0.24 | 0.07 | -0.47*** | 0.27 | -0.17 | -0.07 |
| with < 25,000 people | (0.11) | (0.16) | (0.21) | (0.16) | (0.17) | (0.15) | (0.07) | (0.25) | (0.25) | (0.16) |
| log(population) | 0.06** | 0.12*** | 0.16*** | 0.11*** | -0.06 | -0.18*** | 0.07** | -0.07 | 0.01 | -0.12*** |
| | (0.03) | (0.03) | (0.03) | (0.02) | (0.04) | (0.05) | (0.03) | (0.06) | (0.06) | (0.04) |
| % of Men over 21 | 0.56 | 0.63 | 0.23 | 0.31 | 0.87 | 2.84*** | 0.68 | 2.33*** | 2.16* | 1.25** |
| | (0.40) | (0.48) | (0.57) | (0.33) | (0.58) | (0.70) | (0.47) | (0.89) | (1.30) | (0.63) |
| # Catholic seats per capita | | | | | | | -0.06 | -1.22*** | | -0.33 |
| | | | | | | | (0.08) | (0.39) | | (0.39) |
| # Jewish seats per capita | | | | | | | 65.90*** | -0.80 | -9.98** | -9.05** |
| | | | | | | | (24.04) | (12.28) | (4.47) | (3.86) |
| # seats in all churches per capita | | | | | | | 0.06* | 0.37*** | 0.32*** | 0.26*** |
| | | | | | | | (0.03) | (0.14) | (0.12) | (0.09) |
| Religious Fractionalization | | | | | | | 0.23*** | 0.61*** | 0.48** | 0.45* |
| | | | | | | | (0.08) | (0.18) | (0.20) | (0.24) |
| $R^2$ | | | | | | | 0.27 | 0.33 | 0.26 | 0.22 |
| N = | | | | | | | 392 | 576 | 839 | 1279 |

Table 2.14: Determinants of 19th Century Freemasonry: Panel Data Analysis

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Pooled OLS Estimates | | | Fixed Effect Estimates | | | | |
| Years of Data: | 1850-1900 | 1860-1900 | 1870-1900 | 1850-1900 | 1860-1900 | 1870-1900 | 1850-1890 | 1860-1890 |
| log($Y^M$ per capita) | 0.03 | 0.02 | 0.00 | 0.03 | 0.04* | 0.02 | 0.02 | 0.01 |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.02) | (0.03) | (0.04) | (0.03) |
| log($Y^A$ per capita) | | | 0.07*** | | | 0.06*** | | |
| | | | (0.03) | | | (0.02) | | |
| log($K^M$ per capita) | -0.02 | 0.00 | 0.03 | -0.03 | -0.03 | -0.00 | -0.02 | -0.01 |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.02) | (0.03) | (0.04) | (0.03) |
| log($K^A$ per capita) | 0.06*** | 0.06*** | -0.06** | 0.12*** | 0.10*** | -0.01 | 0.18*** | 0.13*** |
| | (0.02) | (0.01) | (0.02) | (0.04) | (0.02) | (0.02) | (0.06) | (0.05) |
| Proportion of population native born | | 0.44*** | 1.03*** | | -0.54*** | 0.31 | | -0.86** |
| | | (0.09) | (0.12) | | (0.13) | (0.23) | | (0.40) |
| Proportion of population in towns with < 25,000 people | -0.12 | -0.27*** | -0.46*** | -0.94 | -0.52*** | -0.61*** | -0.73 | 0.15 |
| | (0.11) | (0.08) | (0.09) | (0.59) | (0.10) | (0.10) | (0.56) | (0.17) |
| log(population) | -0.00 | 0.08*** | 0.11*** | -0.17 | -0.03 | 0.09*** | -0.31* | -0.49*** |
| | (0.04) | (0.02) | (0.02) | (0.11) | (0.02) | (0.02) | (0.16) | (0.12) |
| % of Men over 21 | 0.90** | 0.54 | 0.63* | -0.15 | -0.57 | -0.25 | -0.19 | -0.38 |
| | (0.41) | (0.33) | (0.33) | (0.60) | (0.39) | (0.37) | (0.54) | (1.40) |
| # Jewish seats per capita | | | | | | | -48.41 | -5.64 |
| | | | | | | | (32.65) | (7.88) |
| # seats in all churches per capita | | | | | | | 0.04 | 0.11 |
| | | | | | | | (0.06) | (0.10) |
| Religious Fractionalization | | | | | | | 0.15 | 0.09 |
| | | | | | | | (0.11) | (0.13) |
| $R^2$ | 0.37 | 0.23 | 0.24 | 0.47 | 0.20 | 0.09 | 0.47 | 0.37 |
| N = | 2552 | 3909 | 3923 | 2552 | 3909 | 3923 | 1650 | 1271 |

Table 2.15: Testing for Autocorrelation in Freemasonry Membership

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Years Included in Panel | 1850-1900 | | 1860-1900 | | 1870-1900 | | 1850-1900‡ | | 1860-1900‡ | |
| Estimator Used: | A-H† | A-B† | A-H† | A-B† | A-H† | A-B† | A-H† | A-B† | A-H† | A-B† |
| Lagged Freemasons | 0.12* | 0.10* | 0.49*** | 0.51*** | 0.31*** | 0.31*** | 0.17 | 0.07 | 0.18 | 0.08 |
| | (0.07) | (0.06) | (0.07) | (0.04) | (0.05) | (0.05) | (0.11) | (0.05) | (0.12) | (0.06) |
| log($Y^M$ per capita) | 0.03 | 0.03 | 0.06* | 0.06* | -0.06* | -0.06* | 0.04 | 0.04 | 0.03 | 0.04 |
| | (0.02) | (0.02) | (0.04) | (0.04) | (0.03) | (0.03) | (0.04) | (0.04) | (0.04) | (0.04) |
| log($Y^A$ per capita) | | | | | 0.13*** | 0.12*** | | | | |
| | | | | | (0.04) | (0.04) | | | | |
| log($K^M$ per capita) | -0.03 | -0.04 | -0.04 | -0.05 | 0.07** | 0.07** | -0.04 | -0.04 | -0.04 | -0.04 |
| | (0.02) | (0.02) | (0.04) | (0.04) | (0.03) | (0.03) | (0.04) | (0.04) | (0.04) | (0.04) |
| log($K^A$ per capita) | 0.07*** | 0.08*** | -0.03 | -0.03 | -0.05 | -0.05 | 0.13*** | 0.14*** | 0.09* | 0.10** |
| | (0.02) | (0.02) | (0.03) | (0.03) | (0.05) | (0.05) | (0.04) | (0.04) | (0.05) | (0.05) |
| Proportion of population native born | | | -0.34 | -0.33 | -0.71 | -0.70 | | | -0.85*** | -0.91*** |
| | | | (0.45) | (0.44) | (0.46) | (0.46) | | | (0.28) | (0.29) |
| Proportion of population in towns with < 25,000 people | 0.09 | 0.09 | -0.32*** | -0.35*** | -0.14 | -0.13 | 0.14 | 0.17 | 0.19 | 0.22 |
| | (0.08) | (0.08) | (0.11) | (0.12) | (0.10) | (0.10) | (0.18) | (0.18) | (0.18) | (0.18) |
| log(population) | -0.45*** | -0.50*** | -0.01 | -0.02 | -0.13* | -0.14** | -0.49*** | -0.51*** | -0.54*** | -0.56*** |
| | (0.08) | (0.09) | (0.06) | (0.06) | (0.07) | (0.07) | (0.08) | (0.09) | (0.08) | (0.09) |
| % of Men over 21 | -2.72*** | -2.70*** | -3.89*** | -3.78*** | -2.41*** | -2.38*** | -2.45*** | -2.21*** | -0.52 | -0.12 |
| | (0.40) | (0.39) | (0.82) | (0.93) | (0.68) | (0.72) | (0.53) | (0.50) | (0.84) | (0.92) |
| # Jewish seats per capita | | | | | | | 0.21 | -2.45 | 0.12 | -2.72 |
| | | | | | | | (8.52) | (8.46) | (8.47) | (8.60) |
| # seats in all churches per capita | | | | | | | -0.00 | 0.00 | 0.00 | 0.01 |
| | | | | | | | (0.08) | (0.08) | (0.08) | (0.08) |
| Religious Fractionalization | | | | | | | -0.01 | 0.02 | -0.02 | 0.01 |
| | | | | | | | (0.13) | (0.12) | (0.13) | (0.12) |
| $R^2$ | 0.23 | | . | | . | | 0.23 | | 0.23 | |
| N = | 1560 | 1560 | 2220 | 2220 | 1848 | 1848 | 724 | 724 | 724 | 724 |

†: A-H refers to the Anderson and Hsiao (1982) estimator using lagged levels of the dependent variable and A-B refers to the Arellano and Bond (1991) estimator using all possible lags of the dependent variable in a GMM framework.
‡: This panel data excludes 1880 as there are no religious data available for this period.

Table 2.16: Freemasonry and Enrollment in the Mid-19th Century (1850)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Enrollment Among Men and Women Ages 5-19 | | | | | |
| | Total | | Public | | Private | |
| log($Y^M$ per capita) | 0.01 | 0.01 | 0.04** | 0.04** | -0.01 | -0.01 |
| | (0.01) | (0.01) | (0.02) | (0.02) | (0.00) | (0.00) |
| log($K^M$ per capita) | -0.01 | -0.01 | -0.04** | -0.04** | 0.01* | 0.01 |
| | (0.01) | (0.01) | (0.02) | (0.02) | (0.00) | (0.00) |
| log($K^A$ per capita) | 0.02*** | 0.02*** | 0.02*** | 0.02*** | 0.00 | -0.00 |
| | (0.01) | (0.01) | (0.00) | (0.01) | (0.00) | (0.00) |
| Proportion of population in towns with < 2,500 people | 0.06 | 0.06 | 0.12 | 0.12 | -0.08*** | -0.09*** |
| | (0.13) | (0.13) | (0.08) | (0.08) | (0.03) | (0.03) |
| Proportion of population in towns with 2,500-25,000 people | 0.04 | 0.03 | 0.02 | 0.01 | 0.04 | 0.03 |
| | (0.14) | (0.14) | (0.10) | (0.10) | (0.05) | (0.05) |
| log(population) | 0.04*** | 0.04*** | 0.05*** | 0.05*** | -0.00 | -0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.00) | (0.00) |
| % of Men over 21 | -0.24* | -0.26* | -0.15 | -0.16 | -0.01 | -0.01 |
| | (0.13) | (0.13) | (0.13) | (0.13) | (0.08) | (0.08) |
| # Masons / Total Population | | 2.49 | | 2.28 | | 1.35 |
| | | (1.67) | | (2.29) | | (0.99) |
| $R^2$ | 0.56 | 0.56 | 0.35 | 0.35 | 0.34 | 0.35 |
| N = | 482 | 482 | 482 | 482 | 482 | 482 |

# Chapter 3

# Efficient estimation of data combination models by the method of auxiliary-to-study tilting (AST) - with Bryan Graham and Cristine Campos de Xavier Pinto

## 3.1 Introduction

Let $Z = (W', X', Y')'$ denote a random vector drawn from some *study population* of interest with distribution function $F_s$. The only prior restriction on $F_s$ is that for some unique $\gamma_0$, and known function $\psi(z, \gamma)$ of the same dimension,

$$\mathbb{E}_s [\psi(Z, \gamma_0)] = 0, \tag{3.1}$$

where $\mathbb{E}_s [\cdot]$ denotes expectations taken with respect to the study population. If a random sample of $Z$ is available, then consistent estimation of $\gamma_0$ (under regularity conditions) is straightforward (e.g., Newey and McFadden, 1994). In this paper we consider estimation of $\gamma_0$ when a random sample of $Z$ from $F_s$ is unavailable. Instead two separate random samples are available. The first is drawn from the *study* population and contains $N_s$ measurements of $(Y, W)$. The second is drawn from an *auxiliary* population (with distribution function $F_a$; $\mathbb{E}_a [\cdot]$ denotes expectations taken with respect to this distribution) and contains $N_a$ measurements of $(X, W)$. While the variable $W$ is common to the two samples, $X$ and $Y$ are not. Hahn (1998) and Chen, Hong and Tarozzi (2008) show that identification of $\gamma_0$ follows if (i) the conditional distributions of $X$ given $W$ in the two populations coincide (although their marginal distributions for $W$ may differ), (ii) the support of $W$ in the auxiliary population is at least as large as that in the study population and (iii) $\psi(z, \gamma_0)$ is separable in the components depending on the 'non-common' variables $Y$ and $X$

$$\psi(Z, \gamma_0) = \psi_s(Y, W, \gamma_0) - \psi_a(X, W, \gamma_0). \tag{3.2}$$

Examples of statistical problems to which the above setup applies include the two sample instrumental variables (TSIV) model of Angrist and Krueger (1992) and Ridder and Moffitt (2007), the average treatment effect on the treated (ATT) estimand from the program evaluation literature (e.g.,Heckman and Robb, 1985; Imbens, 2004), counterfactual earnings/wealth decompositions as in Dinardo, Fortin and Lemieux (1996), Barsky, Bound, Charles and Lupton (2002) and Chernozhukov, Fernandez-Val and Melly (2009), poverty mapping as in Elbers, Lanjouw and Lanjouw (2003) and Tarozzi and Deaton (2007), and models with mismeasured regressors and validation samples (e.g., Caroll and Wand, 1991).

To help fix ideas consider the ATT example. Here $Y$ denotes an individual's potential outcome under active treatment, say earnings given participation in a job training program, $X$ denotes her outcome under control (earnings in the absence of training) and $W$ is a vector of baseline covariates. Available is a random sample of $(Y, W)$ from the population assigned active treatment (i.e., 'the treated'). A separate sample of measurements of $(X, W)$ is drawn from *a* population of controls. The ATT, $\gamma_0 = \mathbb{E}_s [Y - X]$, is given by the solution to

(3.1) with $\psi_s(Y, W, \gamma_0) = Y$ and $\psi_a(X, W, \gamma_0) = X + \gamma_0$.

Methods of estimating the ATT or, more generally, $\gamma_0$ as defined by (3.1), include imputation (e.g., Rubin, 1977, Heckman and Robb, 1985; Hahn, 1998; Chen, Hong and Tarozzi, 2008), a variant of inverse probability weighting (IPW) (e.g., Hirano, Imbens and Ridder, 2003) and matching (e.g., Abadie and Imbens, 2006). Our approach to estimating $\gamma_0$, which we call auxiliary-to-study tilting (AST), differs from each of these methods. In a preliminary step we 'tilt' the auxiliary sample toward the study sample. Let $h(W)$ equal a $M \times 1$ vector of known functions of $W$. Specifically, we use minimum empirical discrepancy (MD) methods, as in Corcoran (1998), to reweight the auxiliary sample such that, after re-weighting, the sample mean of $h(W)$ coincides with its study sample counterpart. For example, we might re-weight to *exactly* balance the auxiliary sample means, variances and pairwise covariances of the elements of $W$ with their (unweighted) study sample counterparts. In this case $h(W)$ would include all own, squares and pairwise cross-products of the elements of $W$.

In a second step we combine sample moments from the study sample with sample moments from the *re-weighted* auxiliary sample to estimate $\gamma_0$ (For example, our estimate of the ATT is the difference in the study sample mean of $Y$ and the MD re-weighted auxiliary sample mean of $X$). We provide conditions under which matching $M$ moments of $W$ across the two samples in this way is sufficient for (i) consistent estimation of $\gamma_0$ and, under further conditions, (ii) efficient. Because we can represent our procedure as a sequential M-estimation problem we are able to use standard results to derive its large sample properties (e.g., Newey and McFadden, 1994).

Conventional MD estimation focuses on efficient estimation of a distribution function. Our focus is different since the auxiliary sample being 'tilted' generally will be incompatible with the imposed constraints (since $F_a(W) \neq F_s(W)$). We show that, in the data combination setting, the choice of discrepancy function and imposed constraints together imply a corresponding parametric model for the the study-to-auxiliary density ratio $f_s(W)/f_a(W)$. Parametric modelling of density ratios is also a component of discriminant analysis. We are therefore able to exploit insights from this literature to guide our selection of the discrepancy function as well as the form of $h(W)$ (e.g., Anderson, 1982).

Chen, Hong and Tarozzi (2008) develop semiparametric efficiency theory for the data combination problem and present a globally efficient estimator (cf., Hahn (1998) for the case of the ATT). Our estimator's variance coincides with their variance bound when two auxiliary parametric restrictions hold in the population (*but these restrictions are not employed in the calculation of the bound* ). Our motivation for taking a 'flexible parametric', as opposed to a fully nonparametric, approach is to develop methods (i) where large sample distribution theory is a good approximation to small sample performance and (ii) that is more concordant with the practice of applied researchers. The implementation of all currently available globally efficient data combination estimators requires high-dimensional nonparametric smoothing. While this does not affect their first order asymptotic properties, it does affect small sample performance (cf., Wang, Linton and Härdle, 2004). That is, in practice the choice of smoothing parameter, for which theory provides limited guidance, may greatly effect one's estimates of $\gamma_0$. The 'flexible parametric' approach taken here is motivated by these curse of dimensionality concerns.

Formally we show that AST's asymptotic variance coincides with the bound of Chen, Hong and Tarozzi (2008, Theorem 1) if (i) our implicit model for $f_s(W)/f_a(W)$, which depends on our choices of discrepancy function and $h(W)$, is correct *and* (ii) $\mathbb{E}_s[\psi_a(X,W,\gamma_0)|W=w] = \varsigma_0 + \Pi_0 h(w)$ for some $K \times 1$ vector $\varsigma_0$ and $K \times M$ matrix $\Pi_0$ (In the ATT example this last condition is equivalent to requiring that the conditional expectation function of the potential outcome under control is linear in $h(w)$). Consistency of our procedure, however, only requires that one of these two assumptions hold. We view this partial robustness as an acceptable consequence of basing estimation on auxiliary parametric assumptions. We note that this type of partial robustness is desirable relative to the extreme fragility of fully parametric methods in this context (cf., Imbens, 2004).

In the next section we formally define the semiparametric data combination problem to which our methods apply and provide some specific examples of problems to which our methods may be useful. Section 3.2 outlines our estimator and characterizes its large sample properties. Section 3.3 uses auxiliary-to-study tilting to estimate the earnings benefit associated with National Supported Work (NSW) demonstration participation as in LaLonde (1986), Heckman and Hotz (1989), and Dehejia and Wahba (1999). We also present the results of a Monte Carlo study which compares our procedure with several leading alternatives. Section 3.4 summarizes, situates our work in the literature, and suggests areas for future research. All proofs are collected in the Appendix.

A formal definition of the data combination model is given by Assumption 3.1.1 below.

**Assumption 3.1.1.** ***Data Combination Model***

*(i)* (IDENTIFICATION) *For some* $\psi(z,\gamma) = \psi_s(y,w,\gamma) - \psi_a(x,w,\gamma)$, *equation (3.1) holds with* $\mathbb{E}_s[\psi(Z,\gamma)] \neq 0$ *for all* $\gamma \neq \gamma_0$, $\gamma \in \mathcal{G} \subset \mathbb{R}^K$, $z \in \mathcal{Z} \subset \mathbb{R}^{\dim(Z)}$.

*(ii)* (CONDITIONAL DISTRIBUTIONAL EQUALITY) $F_s(x|w) = F_a(x|w)$ *for all* $w \in \mathcal{W} \subset \mathbb{R}^{\dim(W)}$, $x \in \mathcal{X} \subset \mathbb{R}^{\dim(X)}$.

*(iii)* (WEAK OVERLAP) *Let* $S_j = \{w : f_j(w) > 0\}$ *for* $j = s, a$, *then* $S_s \subset S_a$.

*(iv)* (MULTINOMIAL SAMPLING) *With probability* $Q_0 \in (\kappa_0, 1-\kappa_0)$ *for* $0 <$

*$\kappa_0 < 1$ we draw a unit at random from $F_s$ and record its realizations of $Y$ and $W$, otherwise we draw a unit at random from $F_a$ and record its realizations of $X$ and $W$. Let $D_i = 1$ if the $i^{th}$ draw ($i = 1, \ldots, N$) corresponds to a study population unit and $D_i = 0$ otherwise.*

The first part of Assumption 3.1.1 implies global identifiability of the complete data model. The second part implies that the distribution of $(X, W)$ in the two populations differ only in terms of their marginal distributions for the 'always observed' variable, $W$. The third part ensures that, in large samples, for each unit in the study sample there will be 'matching' units with similar values of $W$ in the auxiliary sample. The fourth part of Assumption 3.1.1 allows us to treat the *merged sample*

$$\left\{ (D_i, W_i,' (1 - D_i) X_i', D_i Y_i')' \right\}_{i=1}^{N},$$

'as if' it were a random one from a hypothetical *merged population* with distribution function $F$ (let $\mathbb{E}[\cdot]$ denote expectations taken with respect to this distribution). The semiparametric data combination model is typically defined by specifying properties of the merged population (e.g., Chen, Hong and Tarozzi, 2008). We prefer the formulation given above because it emphasizes that the problem is fundamentally one of data combination.

The sampling distribution induced by the multinomial scheme, $F$, has density

$$f(z, d) = Q_0^d (1 - Q_0)^{1-d} f_s(z)^d f_a(z)^{1-d},$$

such that $f(z|\, d = 1) = f_s(z)$ and $f(z|\, d = 0) = f_a(z)$. Now consider the conditional probability given $W = w$ that a unit in the merged sample corresponds to a draw from the study population. Let $\mathbb{E}[D|\, W = w] = p_0(w)$ denote this 'propensity score', by Bayes' Law we can define a relationship between the study and auxiliary densities of $W$ in terms of $p_0(w)$

$$f_s(w) = f_a(w) \left\{ \frac{1 - Q_0}{Q_0} \frac{p_0(w)}{1 - p_0(w)} \right\}. \tag{3.3}$$

Under the merged population formulation of the problem it is clear that part one of Assumption 3.1.1 corresponds to requiring that $\mathbb{E}[\psi(Z, \gamma_0)|\, D = 1] = 0$, part two to a conditional independence restriction on the merged population distribution function of $F(x|\, w, d = 1) = F(x|\, w, d = 0)$, and part three to assuming that $p_0(w)$ is bounded away from one.

Identification of $\gamma_0$ follows from, using parts two and three of Assumption

3.1.1 and Equation (3.3), the equality

$$\begin{aligned}\mathbb{E}_s\left[\psi\left(Z,\gamma\right)\right] &= \mathbb{E}\left[\psi_s\left(Y,W,\gamma\right)\right| D=1] - \mathbb{E}\left[\psi_a\left(X,W,\gamma\right)\right| D=1] \qquad (3.4)\\ &= \mathbb{E}\left[\frac{D}{Q_0}\psi_s\left(Y,W,\gamma\right)\right] - \mathbb{E}\left[\frac{1-D}{Q_0}\frac{p_0\left(W\right)}{1-p_0\left(W\right)}\psi_a\left(X,W,\gamma\right)\right],\end{aligned}$$

which is, by part one of Assumption 3.1.1, uniquely zero at $\gamma = \gamma_0$.

Semiparametric efficiency theory for the model defined by Assumption 3.1.1 is well-developed. We present the key results here as we use them to verify the efficiency of AST below. The maximal asymptotic precision with which $\gamma_0$ may be estimated, derived by Hahn (1998) and Chen, Hong and Tarozzi (2008), is given by the inverse of

$$\mathcal{I}\left(\gamma_0\right) = \mathbb{E}\left[\frac{p_0\left(W\right)}{Q_0}\Gamma_0\left(W\right)\right]' \mathbb{E}\left[\Phi_0\left(W\right)\right]^{-1} \mathbb{E}\left[\frac{p_0\left(W\right)}{Q_0}\Gamma_0\left(W\right)\right], \qquad (3.5)$$

with $\Gamma_0\left(w\right) = \mathbb{E}\left[\partial\psi\left(Z,\gamma_0\right)/\partial\gamma'\right| W=w]$ and

$$\begin{aligned}\Phi_0\left(w\right) = \left\{\frac{p_0\left(w\right)}{Q_0}\right\}^2 &\left\{\frac{\Sigma_s\left(w;\gamma_0\right)}{p_0\left(w\right)} + \frac{\Sigma_a\left(w;\gamma_0\right)}{1-p_0\left(w\right)}\right. \qquad (3.6)\\ &\left.+\frac{1}{p_0\left(w\right)}\left[q_s\left(w;\gamma_0\right)-q_a\left(w;\gamma_0\right)\right]\left[q_s\left(w;\gamma_0\right)-q_a\left(w;\gamma_0\right)\right]'\right\},\end{aligned}$$

where $q_s\left(w;\gamma_0\right) = \mathbb{E}\left[\psi_s\left(Y,W,\gamma_0\right)\right| W=w]$, $q_a\left(w;\gamma_0\right) = \mathbb{E}\left[\psi_a\left(X,W,\gamma_0\right)\right| W=w]$, $\Sigma_s\left(w;\gamma_0\right) = \mathbb{V}\left(\psi_s\left(Y,W,\gamma_0\right)\right| W=w)$ and $\Sigma_a\left(w;\gamma_0\right) = \mathbb{V}\left(\psi_a\left(X,W,\gamma_0\right)\right| W=w)$.

To give some idea of the range of problems to which our methods apply, we outline two examples below (in addition to the program evaluation example discussed in the introduction). Additional examples are described in Chen, Hong and Tarozzi (2008), Ridder and Moffitt (2007) and a supplemental Web Appendix.

**Two sample instrumental variables (TSIV) model:** Ridder and Moffitt (2007) consider two sample instrumental variables (TSIV) models of the form

$$\mathbb{E}_s\left[\left\{f\left(Y;\gamma\right)-g\left(X,W_1;\gamma\right)\right\}e\left(W\right)\right]=0,$$

with $W = \left(W_0', W_1'\right)'$. The firstsample consists of measurements of $(Y,W)$ and the second of $(X,W)$. They assume that both samples are random ones from the study population (i.e., the samples are 'compatible'). This corresponds to augmenting Assumption 3.1.1 with the additional requirement that $F_s\left(w\right) = F_a\left(w\right)$. The TSIV model is of the form required by (3.2) with $\psi_s\left(y,w,\gamma\right) = f\left(Y;\gamma\right)e\left(W\right)$ and $\psi_a\left(x,w,\gamma\right) = g\left(X,W_1;\gamma\right)e\left(W\right)$. When $e\left(W\right) = W$, $f\left(Y;\gamma\right) = Y$ and $g\left(X,W_1;\gamma\right) = X'\alpha + W_1'\beta$ with $\gamma_0 = \left(\alpha_0,\beta_0'\right)'$ we have the linear model analyzed by Angrist and Krueger (1992). Ridder and Moffitt (2007) show how one may estimate the Mixed Proportional Hazard (MPH) model under this setup, while Ichimura and Martinez-Sanchis (2004) discuss binary choice models.

In applications of the TSIV model it is often found that the sample moments of the common variables $W$ differ significantly across the two datasets being combined (e.g., Björkland and Jäntti, 1997). This suggests that full compatibility may fail in practice (i.e., $F_s(w) \neq F_a(w)$). The AST estimator presented below does not require full compatibility and is generally more efficient than the one proposed by Angrist and Krueger (1992) (compare Theorem 3.2.1 below with Angrist and Krueger (1992, p. 331) or Moffitt and Ridder (2007, p. 5505)).

**Poverty mapping:** Let $X$ be an indicator denoting whether a household's total outlay falls below a poverty line and $W$ a vector of household characteristics. We seek to estimate the poverty rate in a specific study municipality as in Elbers, Lanjouw and Lanjouw (2003) and Tarozzi and Deaton (2007). Available is a random sample of $N_s$ observations of $W$ from this municipality; however, no poverty measurements are available in this sample. Also available is a random sample of size $N_a$ of both $X$ and $W$ from the entire country. Our estimand is $\gamma_0 = \mathbb{E}_s[X]$ which corresponds to setting $\psi_s(Y, W, \gamma) = 0$ and $\psi_a(X, W, \gamma) = X - \gamma$. In this example part two of

Assumption 3.1.1 implies that the conditional probability of begin poor given $W = w$ is the same in the entire country as it is in the specific municipality of interest.

## 3.2 Estimation

In this section we describe our procedure and characterize its large sample properties. Full implementation details may be found in a supplemental Web Appendix. Before describing the general case we start by considering, in sequence, the (i) population and (ii) asymptotic sampling properties of auxiliary-to-study tilting (AST) in a very simple setting. This setting corresponds to our last motivating example: poverty mapping. We identify $\gamma_0$ by

$$\gamma_0 = \int \int x \frac{1 - Q_0}{Q_0} \frac{p_0(w)}{1 - p_0(w)} f_a(x, w) \, \mathrm{d}x \mathrm{d}w.$$

**Population problem:** The population analog of the AST estimate of $\gamma_0$ is the mean of $X$ under the 'tilted' distribution $F_*(x, w)$ :

$$\mathbb{E}_{F_*}[X] = \int x f_*(x, w) \, \mathrm{d}x \mathrm{d}w. \tag{3.7}$$

To describe the construction of the tilted distribution function let $h(W)$ be an $M \times 1$ vector of functions that have mean $\zeta_s$ in the study population. A leading form for $h(W)$ is $h(W) = \left(W, W^2, \ldots, W^M\right)'$, which implies that $\zeta_s$ equals $W's$ first $M$ uncentered study population moments (which we assume exist). Note that $Q_0$, $p_0(w)$ and $\zeta_s$ are all asymptotically identified by the sampling process since $(D, W')'$ is observed for all units in the merged sample.

The *auxiliary-to-study tilt* (AST) of $f_a(x, w)$ is given by

$$f_*(x, w) = f_a(x, w) \left\{ \frac{1 - Q_0}{Q_0} \frac{G\left(\alpha_* + h(W)'\beta_*\right)}{1 - G\left(\alpha_* + h(W)'\beta_*\right)} \right\},$$

where $\alpha_*$ and $\beta_*$ are the solutions to the concave programming problem

$$\max_{\alpha,\beta} \left\{ \alpha + \zeta_s'\beta - \mathbb{E}_a\left[\varphi^+\left(\alpha + h(W)'\beta; Q_0\right)\right] \right\}, \tag{3.8}$$

with $\varphi^+(v; Q)$ given by

$$\varphi^+(v, Q) = v \frac{1 - Q}{Q} \frac{G(v)}{1 - G(v)} + \int_{\frac{1-Q}{Q} \frac{G(v)}{1-G(v)}}^{a} G^{-1}\left(\frac{t}{t + (1 - Q)/Q}\right) \mathrm{d}t \Bigg\}, \tag{3.9}$$

and $G(\cdot)$ is a strictly increasing, differentiable and continuous function mapping the real line onto the unit interval and $a$ is some arbitrary constant.

The $1 + M$ first order conditions for (3.8) are

$$\begin{aligned} \int \frac{1 - Q_0}{Q_0} \frac{G\left(\alpha_* + h(w)'\beta_*\right)}{1 - G\left(\alpha_* + h(w)'\beta_*\right)} f_a(w)\, \mathrm{d}w &= 1 \\ \int h(W) \frac{1 - Q_0}{Q_0} \frac{G\left(\alpha_* + h(w)'\beta_*\right)}{1 - G\left(\alpha_* + h(w)'\beta_*\right)} f_a(w)\, \mathrm{d}w &= \zeta_s. \end{aligned} \tag{3.10}$$

Equation (3.10) show that the auxiliary-to-study tilt is chosen so that it integrates to one and shares (at least) $M$ moments of $W$ with $F_s$. This is the 'exact balancing' property of AST.

Our main identification result is to show that the expectation of $X$ under the tilted distribution coincides with its study population expectation (i.e, $\mathbb{E}_{F_*}[X] = \mathbb{E}_s[X] = \gamma_0$) if *at least one of* two auxiliary parametric restrictions holds:

$$\frac{f_s(w)}{f_a(w)} = \frac{1 - Q_0}{Q_0} \frac{G\left(\alpha_0 + h(w)'\beta_0\right)}{1 - G\left(\alpha_0 + h(w)'\beta_0\right)} \tag{3.11}$$

for some scalar $\alpha_0$ and $M \times 1$ vector $\beta_0$, or

$$\mathbb{E}_s[X | W = w] = \varsigma_0 + \Pi_0 h(w) \tag{3.12}$$

for $\varsigma_0$ and some $1 \times M$ vector $\Pi_0$. This is our partial or 'double' robustness result.

Equality of $\mathbb{E}_{F_*}[X]$ and $\gamma_0$ under (3.11) follows from global concavity of $\varphi^+(v; Q)$ and the consequent equalities $\alpha_* = \alpha_0$ and $\beta_* = \beta_0$ (which implies that $f_*(x, w)$ and $f_s(x, w)$ coincide). Equality of $\mathbb{E}_{F_*}[X]$ and $\gamma_0$ under the second condition is less obvious and an important quality of auxiliary-to-study tilting. By iterated expectations and part two of Assumption 3.1.1 we have from (3.7)

$$\mathbb{E}_{F_*}[X] = \int \mathbb{E}_s[X|W = w] \left\{ \frac{1 - Q_0}{Q_0} \frac{G(\alpha_* + h(w)'\beta_*)}{1 - G(\alpha_* + h(w)'\beta_*)} \right\} f_a(w)\,\mathrm{d}w.$$

Substituting in $\varsigma_0 + \Pi_0 h(w)$ for $\mathbb{E}_s[X|W = w]$ gives

$$\begin{aligned}
&\mathbb{E}_{F_*}[X] = \varsigma_0 \int \left\{ \frac{1 - Q_0}{Q_0} \frac{G(\alpha_* + h(w)'\beta_*)}{1 - G(\alpha_* + h(w)'\beta_*)} \right\} f_a(w)\,\mathrm{d}w \\
&+ \Pi_0 \int h(w) \left\{ \frac{1 - Q_0}{Q_0} \frac{G(\alpha_* + h(w)'\beta_*)}{1 - G(\alpha_* + h(w)'\beta_*)} \right\} f_a(w)\,\mathrm{d}w = \varsigma_0 + \Pi_0 \zeta_s = \mathbb{E}_s[X],
\end{aligned}$$

where the second equality follows from the first order conditions for $(\alpha_*, \beta_*)$ given by (3.10) and the third from iterated expectations. The mean of $X$ with respect to the tilted auxiliary distribution will equal its study population mean if the two distributions are sufficiently similar. For example, if $W$ is a scalar and $\mathbb{E}_s[X|W = w]$ is quadratic in $w$, then requiring that the auxiliary-to-study tilt of $F_a$ shares the same mean and variance for $W$ as $F_s$ ensures that $\mathbb{E}_{F_*}[X] = \gamma_0$.

**Sample problem:** When both of the auxiliary restrictions (3.11) and (3.12) hold we can show, by iterated expectations, that (3.10) and (3.7) are equivalent to the pair of moment restrictions

$$\mathbb{E}\left[ \left\{ (1 - D) \frac{G(\alpha_0 + h(W)'\beta_0)}{1 - G(\alpha_0 + h(W)'\beta_0)} - D \right\} \begin{pmatrix} 1 \\ h(W) \end{pmatrix} \right] = 0 \quad (3.13)$$

$$\mathbb{E}\left[ (1 - D) \frac{G(\alpha_0 + h(W)'\beta_0)}{1 - G(\alpha_0 + h(W)'\beta_0)} (X - \gamma_0) \right] = 0. \quad (3.14)$$

Our AST estimator is simply the sequential method-of-moments estimate of $\gamma_0$ based on these two restrictions. In particular, $\widehat{\gamma}$ solves

$$\frac{1}{N} \sum_{i=1}^{N} (1 - D_i) \frac{G(\widehat{\alpha} + h(W_i)'\widehat{\beta})}{1 - G(\widehat{\alpha} + h(W_i)'\widehat{\beta})} (X_i - \widehat{\gamma}) = 0, \quad (3.15)$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are first step estimates of $\alpha_0$ and $\beta_0$ defined by the equality

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1-D_i}{\widehat{Q}}\frac{G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}{1-G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}\begin{pmatrix}1\\ h(W_i)\end{pmatrix}=\begin{pmatrix}1\\ \widehat{\zeta}_s\end{pmatrix}, \qquad (3.16)$$

with $\widehat{Q}=N_s/N$ and $\widehat{\zeta}_s=\sum_{i=1}^{N}D_ih(W_i)/N_s$. Equation (3.16) shows that $\widehat{\alpha}$ and $\widehat{\beta}$ are chosen so that the tilting weights, $\frac{1}{N}\frac{1-D_i}{\widehat{Q}}\frac{G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}{1-G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}$, sum to one and the tilted mean of $h(W_i)$ equals its study sample (unweighted) mean. These are the sample analogs of (3.10) above.

An implication of (3.16) is that the tilted mean of $\varsigma_0+\Pi_0h(W_i)-\gamma_0$ in the auxiliary sample equals its unweighted study sample mean:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)\,G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}{1-G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}\left[\varsigma_0+\Pi_0h(W_i)-\gamma_0\right]=\frac{1}{N}\sum_{i=1}^{N}D_i\left[\varsigma_0+\Pi_0h(W_i)-\gamma_0\right].$$

Using this exact balancing implication of (3.16) we can add and subtract terms to (3.15) to yield

$$\begin{aligned}0 \;=\;& \frac{1}{N}\sum_{i=1}^{N}(1-D_i)\frac{G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}{1-G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}\left(X_i-\gamma_0-(\widehat{\gamma}-\gamma_0)\right)\\ &-\frac{1}{N}\sum_{i=1}^{N}(1-D_i)\frac{G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}{1-G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}\left[\varsigma_0+\Pi_0h(W_i)-\gamma_0\right]\\ &+\frac{1}{N}\sum_{i=1}^{N}D_i\left[\varsigma_0+\Pi_0h(W_i)-\gamma_0\right].\end{aligned}$$

For this model we have $\psi_s(Y,W,\gamma)=0$ and $\psi_a(X,W,\gamma)=X-\gamma$ and, under (3.12), $q_a(w;\gamma_0)=\varsigma_0+\Pi_0h(w)-\gamma_0$. Solving for $\sqrt{N}(\widehat{\gamma}-\gamma_0)$ and making these substitutions therefore gives

$$\begin{aligned}\sqrt{N}(\widehat{\gamma}-\gamma_0)=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1-D_i}{\widehat{Q}}\frac{G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}{1-G(\widehat{\alpha}+h(W_i)'\widehat{\beta})}\left\{\psi_a(X_i,W_i,\gamma_0)-q_a(W_i;\gamma_0)\right\}\\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{D_i}{\widehat{Q}}q_a(W_i;\gamma_0).\end{aligned}$$

Finally, a mean value expansion in $\widehat{Q}$, $\widehat{\alpha}$ and $\widehat{\beta}$ about $Q_0$, $\alpha_0$ and $\beta_0$ yields the asymptotically linear representation

$$\sqrt{N}\left(\widehat{\gamma}-\gamma_0\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-D_i}{Q_0}\frac{p_0\left(W_i\right)}{1-p_0\left(W_i\right)}\left\{\psi_a\left(X_i,W_i,\gamma_0\right)-q_a\left(W_i;\gamma_0\right)\right\}\right.$$
$$\left.+\frac{D_i}{Q_0}q_a\left(W_i;\gamma_0\right)\right\}+o_p\left(1\right).$$

The variance of the term in $\{\cdot\}$ equals the inverse of the semiparametric information bound of Hahn (1998) and Chen, Hong and Tarozzi (2008), given in (3.5) above.

### 3.2.1 Estimation in the general case

As noted in the introduction, AST is usefully viewed as a minimum empirical discrepancy (MD) estimator. The 'tilting' of the auxiliary sample is achieved by solving a minimum empirical discrepancy problem: we choose a reweighting of the auxiliary sample that is as 'close as possible' to its empirical measure subject to the restriction that the resulting estimated measure shares at least $M$ moments of $W$ with the study sample. 'Close as possible' is given content by choosing a specific discrepancy function.

We work with contrast functions of the form

$$\varphi\left(v;Q\right)\overset{def}{\equiv}\begin{cases}-\frac{v}{k(Q)}G^{-1}\left(Q\right)-\frac{1}{k(Q)}\int_v^a G^{-1}\left(\frac{t}{t+(1-Q)/Q}\right)\mathrm{d}t & v>0\\ +\infty & v\leq 0\end{cases},\tag{3.17}$$

where the function $G\left(\cdot\right)$ is strictly increasing, differentiable and maps into the unit interval with $\lim_{v\rightarrow-\infty}G\left(v\right)=0$ and $\lim_{v\rightarrow\infty}G\left(v\right)=1$. We also require that, $G_1\left(v\right)=\partial G\left(v\right)/\partial v$, is symmetric about zero. Observe that (3.17) is convex, differentiable and attains its minimum at $v=1$. The term $k\left(Q\right)=Q\left(1-Q\right)/G_1\left(G^{-1}(Q)\right)$ is a normalizing constant; its presence facilitates asymptotic analysis.

The tilt of the auxiliary sample is given by the solution to

$$\min_{\pi_{N_s+1},\ldots,\pi_N}\quad\frac{1}{N_a}\sum_{i=N_s+1}^{N}\varphi(N_a\pi_i;\widehat{Q}),\quad s.t.\quad\sum_{i=N_s+1}^{N}\pi_i=1,\quad\sum_{i=N_s+1}^{N}\pi_i h(W_i,\widehat{\zeta}_s)=0,\tag{3.18}$$

with $h(W_i,\zeta_s)=h(W_i)-\zeta_s$. We assume, without loss of generality, that the first $N_s$ units in the merged sample corresponds to the study sample, with the remainder being the auxiliary sample. Let $\widehat{\rho}=(\widehat{Q},\widehat{\zeta}_s')'$, $\mathcal{L}\left(\underline{\pi},\eta,\lambda;\widehat{\rho}\right)$ be the Lagrangian associated with (3.18) (with $\eta$ the multiplier associated with the adding up constraint and $\lambda$ the $M$ multipliers associated with the moment balancing constraints); the

MD probability weights take the form

$$\widehat{\pi}_i = \frac{1}{N_a}\frac{1-\widehat{Q}}{\widehat{Q}}\frac{G(k(\widehat{Q})t(W_i,\widehat{\zeta}_s)'\widehat{\delta}+G^{-1}(\widehat{Q}))}{1-G(k(\widehat{Q})t(W_i,\widehat{\zeta}_s)'\widehat{\delta}+G^{-1}(\widehat{Q}))}, \qquad i = N_s+1,\ldots,N, \quad (3.19)$$

for $t(W,\zeta) = \left(1, h\left(W,\zeta\right)'\right)'$ and $\widehat{\delta} = (\widehat{\eta},\widehat{\lambda}')'$. The vector of Lagrange multipliers solve

$$\frac{1}{N_a}\frac{1-\widehat{Q}}{\widehat{Q}}\sum_{i=N_s+1}^{N}\frac{G(k(\widehat{Q})t(W_i,\widehat{\zeta}_s)'\widehat{\delta}+G^{-1}(\widehat{Q}))}{1-G(k(\widehat{Q})t(W_i,\widehat{\zeta}_s)'\widehat{\delta}+G^{-1}(\widehat{Q}))}t(X_i,\widehat{\zeta}_s) - t_0 = 0, \qquad (3.20)$$

for $t_0 = (1, \underline{0}')'$.

Inspection of (3.20) indicates the tilt, $F_{\widehat{\pi}}\left(x,w\right) = \sum_{i=N_s+1}^{N}\widehat{\pi}_i\mathbf{1}\left(X_i \leq x, W_i \leq w\right)$, is chosen to be as close as possible to the empirical measure of auxiliary sample while simultaneously satisfying the balancing restrictions $\int h\left(w\right)\mathrm{d}F_{\widehat{\pi}}\left(x,w\right) = \int h\left(w\right)\mathrm{d}F_{N_s}\left(w\right)$, as well as the 'adding-up' condition $\int \mathrm{d}F_{\widehat{\pi}}\left(x,w\right) = 1$.

In the final step, the AST estimate of $\gamma_0$ is given by the solution to

$$0 = \frac{1}{N_s}\sum_{i=1}^{N_s}\psi_s\left(Y_i,W_i,\widehat{\gamma}\right) - \sum_{i=N_s+1}^{N}\widehat{\pi}_i\psi_a\left(X_i,W_i,\widehat{\gamma}\right).$$

The MD formulation's value is primarily pedagogical. For estimation, as well as to characterize large sample properties, a sequential method of moments formulation is more convenient. In step one $\widehat{Q}$ and $\widehat{\zeta}_s$ are chosen to solve

$$\frac{1}{N}\sum_{i=1}^{N}m_1(Z_i,\widehat{\rho}) = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} D_i - \widehat{Q} \\ \frac{D_i}{\widehat{Q}}(h\left(W_i\right) - \widehat{\zeta}_s)\end{pmatrix} = 0. \qquad (3.21)$$

In step two the tilt of the auxiliary sample is computed. To compute the tilt it is convenient to work with the dual problem (cf., Kitamura, 2007). The Fenchel duality theorem implies that $\widehat{\delta}$ (when it exists) is equal to the solution to

$$\max_{\delta}\left\{t_0'\delta + \frac{1}{N_a}\sum_{i=N_s+1}^{N}\varphi^{+}(t(W_i,\widehat{\zeta}_s)'\delta,\widehat{Q})\right\},$$

where $\varphi^{+}(v,Q)$ is the negative of the Fenchel conjugate of $\varphi(v,Q)$:

$$\varphi^{+}(v,Q) = -\frac{1}{k(Q)}\left\{\left[k(Q)v + G^{-1}(Q)\right]\frac{1-Q}{Q}\frac{G(k(Q)v+G^{-1}(Q))}{1-G(k(Q)v+G^{-1}(Q))}\right\} + \int_{\frac{1-Q}{Q}\frac{G(k(Q)v+G^{-1}(Q))}{1-G(k(Q)v+G^{-1}(Q))}}^{a} G^{-1}\left(\frac{t}{t+(1-Q)/Q}\right)\mathrm{d}t\Bigg\}.$$

The first order condition to this problem implies that $\widehat{\delta}$ solves

$$\frac{1}{N}\sum_{i=1}^{N} m_2(Z_i,\widehat{\rho},\widehat{\delta}) = \frac{1}{N}\sum_{i=1}^{N}\left(t_0 + \frac{1-D_i}{1-\widehat{Q}}\varphi_1^{+}(t(W_i,\widehat{\zeta}_s)'\widehat{\delta},\widehat{Q})t(W_i,\widehat{\zeta}_s)\right) = 0, \tag{3.22}$$

where

$$\varphi_1^{+}(t(W_i,\widehat{\zeta}_s)'\widehat{\delta},\widehat{Q}) = -\frac{1-\widehat{Q}}{\widehat{Q}}\frac{G(k(\widehat{Q})t(W_i,\widehat{\zeta}_s)'\widehat{\delta}+G^{-1}(\widehat{Q}))}{1-G(k(\widehat{Q})t(W_i,\widehat{\zeta}_s)'\widehat{\delta}+G^{-1}(\widehat{Q}))}.$$

The probability weights can then be recovered by $\widehat{\pi}_i = -\varphi_1^{+}(t(W_i,\widehat{\zeta}_s)'\widehat{\delta})/N_a$. Finally $\widehat{\gamma}$ solves

$$\frac{1}{N}\sum_{i=1}^{N} m_3(Z_i,\widehat{\rho},\widehat{\delta},\widehat{\gamma}) = -\frac{1}{N}\left\{\sum_{i=1}^{N} -\frac{D_i}{\widehat{Q}}\psi_s(Y_i,W_i,\widehat{\gamma}) - \frac{(1-D_i)\varphi_1^{+}(t(W_i,\widehat{\zeta}_s)'\widehat{\delta},\widehat{Q})}{1-\widehat{Q}}\psi_a(X_i,W_i,\widehat{\gamma})\right\}. \tag{3.23}$$

For a given choice of $h(W)$, different choices of $G(v)$ – and hence $\varphi^{+}(v,Q)$ – will lead to different estimates of $\gamma$. In the empirical application we work with the logit specification $G(v) = \exp[v]/[1+\exp[v]]$, and suspect that others will find this choice useful in practice as well. For that choice it is easy to show that $\varphi^{+}(v,Q) = -\exp[v]$, which is the criterion function associated with the exponential tilting (ET) estimator (e.g., Kitamura, 2007).

In order to discuss the large sample properties of $\widehat{\gamma}$ in the general data combination problem it is helpful to introduce the following additional assumptions.

**Assumption 3.2.1.** (Density Ratio) *There is a unique* $(\alpha_0,\beta_0) \in \mathcal{A}\times\mathcal{B} \subset \mathbb{R}^{1+M}$ *such that*

$$\frac{f_s(w)}{f_a(w)} = \frac{1-Q_0}{Q_0}\frac{G(\alpha_0+h(w)'\beta_0)}{1-G(\alpha_0+h(w)'\beta_0)}$$

*for all* $w \in \mathcal{W}$.

**Assumption 3.2.2.** (Moment CEF) *For* $q_a(w;\gamma_0) = \mathbb{E}[\psi_a(X, W, \gamma_0)|W = w]$ *and some* $(\varsigma_0, \Pi_0) \in \mathcal{S} \times \mathcal{P} \subset \mathbb{R}^K \times \mathbb{R}^{KM}$ :

$$q_a(w;\gamma_0) = \varsigma_0 + \Pi_0 h(w)$$

*for all* $w \in \mathcal{W}$.

Assumption 3.2.1 simply states that the working model for the density ratio implicit in the minimum empirical discrepancy procedure is correctly specified. For the logistic case Assumption 3.2.1 implies

$$\ln\left\{\frac{f_s(w)}{f_a(w)}\right\} = \ln\left(\frac{1-Q_0}{Q_0}\right) + \alpha_0 + h(w)'\beta_0, \tag{3.24}$$

which is identical to the type of restriction made in logistic discriminant analysis (e.g., Anderson, 1982). If both $f_s(w)$ and $f_a(w)$ are multivariate normal, then (3.24) will hold if $h(W)$ includes all own, square and pairwise cross products of the individual elements of $W$. Assumption 3.2.2 is less standard. Its precise content depends on the form of $\psi_a(X, W, \gamma_0)$ and hence its evaluation is application specific. We discuss it in the context of our ATT example below.

We provide distribution theory for $\widehat{\gamma}$ appropriate for two cases: (i) either Assumption 3.2.1 or 3.2.2, but not both, is true and (ii) both are true. We could also consider the case where neither of these two assumptions hold. This would characterize the large sample properties of $\widehat{\gamma}$ under misspecification (cf., Hellerstein and Imbens, 1999). While we do not pursue this case here, we note that the asymptotic variance estimator suggested in the Appendix can also be used to conduct valid inference about the probability limit of $\widehat{\gamma}$ under misspecification.

**Theorem 3.2.1.** (Large Sample Properties) *Suppose Assumption 3.1.1 holds:*

*(i) If either Assumption 3.2.1 or 3.2.2, and additional regularity conditions hold, then* $\widehat{\gamma} \xrightarrow{p} \gamma_0$ *and* $\sqrt{N}(\widehat{\gamma} - \gamma_0) \xrightarrow{D} \mathcal{N}(0, \Upsilon_0)$ *(with the exact form of* $\Upsilon_0$ *depending on whether Assumption 3.2.1 or 3.2.2holds).*

*(ii) If both Assumptions 3.2.1 or 3.2.2 hold, then* $\Upsilon_0 = \mathcal{I}(\gamma_0)^{-1}$ *where* $\mathcal{I}(\gamma_0)$ *is defined by (3.5).*

*Proof.* See Appendix C.1.

□

Assumptions 3.2.1 and 3.2.2 provide insight in how to choose $G(v)$ and $h(w)$ in order to maximize the chance that the Assumptions of Theorem 3.2.1 are satisfied.

Satisfying Assumption 3.2.1 requires prior knowledge on the nature of any differences in the distribution of $W$ in the study and auxiliary populations. This prior information can be expressed either in the form of a parametric model for $f_s(w)/f_a(w)$, as in the statement of Assumption 3.2.1, or, equivalently, in terms of a merged population 'propensity score' restriction of $p_0(w) = G(\alpha_0 + h(w)'\beta_0)$. We suspect that in most data combination setting such prior information is likely to be either very sharp or rather diffuse. An example of sharp prior information is when the auxiliary sample is a stratified random sample from the study population. In such cases $h(w)$ can consist of a vector of dummy variables for each of the sampling strata. When prior information is diffuse a simple default would be to have $h(W)$ include all own and cross product terms of $W$ as well as squared terms for non-binary elements.

Choosing $h(W)$ to satisfy Assumption 3.2.2 involves application-specific considerations. For the ATT example it corresponds to assuming that the conditional mean of the potential outcome under control treatment is linear in $h(w)$. While Assumption 3.2.2 may appear somewhat abstract, it is precisely the type of restriction required to implement widely-used parametric imputation procedures (e.g., Rubin, 1977).

## 3.3 Application and Monte Carlo experiments

LaLonde (1986) compared a variety of non-experimental program evaluation estimators with an experimental benchmark using data from the National Supported Work (NSW) demonstration and the Current Population Survey (CPS). Heckman and Hotz (1989) and Dehejia and Wahba (1999) revisited LaLonde's work using alternative non-experimental program evaluation estimators. In this section we examine LaLonde's data once again. We use AST to estimate the effect of NSW participation on the post-treatment earnings of participants (i.e., the average treatment effect on the treated (ATT)). Dehejia and Wahba (1999) provide full details on the NSW demonstration and the particular data extracts we employ.

The evaluation dataset consists of 297 NSW participants and 425 controls. Our study sample consists of the subset of the 297 NSW participants who are (i) black and (ii) have two years of pre-intervention earnings information ($N_s = 156$). These two restrictions are made to, respectively, (i) improve covariate overlap with CPS sample, and (ii) to make the assumption of conditional distributional equality more plausible. Our ATT is for the associated subpopulation.

Column 1 of Table 3.1 reports summary statistics for this subsample. An auxiliary sample of black non-participants was drawn from the CPS ($N_s = 1,176$, see Column 2). The Column 1 and 2 samples form the basis of our non-experimental estimates of NSW participation on post-treatment earnings for blacks. Column 3 reports the difference in the mean characteristics across the NSW participant and CPS control sample; the two samples differ substantially in terms of pre-intervention characteristics.

Using logistic regression we fitted a model for the propensity score using the merged sample of NSW participants and CPS controls. The propensity score took a logit form with each of the twelve pre-treatment characteristics listed in Table 3.1 entering linearly. The associated difference in pre-treatment characteristics after inverse probability reweighting the CPS controls is reported in Column 4 of Table 3.1 (the 'inverse probability weights' are of the form suggested by Imbens (2004, p. 17) for estimation of the ATT). Inverse probability weighting does lead to considerably more balance in pre-treatment characteristics across NSW participants and CPS controls. Nevertheless IPW has difficulty balancing the mean, variance and covariance of pre-intervention earnings (e.g., even after reweighting the CPS controls' pre-intervention earnings are about 10 percent higher in 1974 and 1975).

By construction, the AST estimator, reweights the CPS data to exactly match the sample moments of the NSW study sample. Table 3.1 verifies this claim, showing that after tilting the auxiliary sample the differences in the two sets of sample moments are identically zero (Column 5). In implementing AST we assumed that $G(v)$ was logistic with $h(W)$ including each function of the covariates listed in Table 3.1.

In Table 3.2 we report estimates of the average effect of participation on post-intervention earnings (i.e., the ATT). Since the NSW data comes from a random assignment experiment, a difference in post-intervention earnings across black NSW participants and non-participants (with two years of pre-intervention earnings data) *in the evaluation dataset* can serve as a 'benchmark' against which alternative, non-experimental, estimators can be compared. This difference is reported in Column 1 of Table 3.2. The experimental ATT estimate suggests statistically significant and economically sizable wage gains for black male participants.

The next column of Table 3.2 reports a parametric imputation (PI) estimate of the ATT as in Rubin (1977) (cf., Imbens, 2004). This estimate is based on a model for the conditional expectation of potential earnings under control that is linear in each function of the covariates listed in Table 3.1. Column 3 reports IPW estimates using the logit propensity score described above (cf., Hirano, Imbens and Ridder, 2003). Column 4 lists our AST estimates. All of the non-experimental estimators reproduce the experimental benchmark up to sampling error and of similar magnitude. Each observational point estimate, however, is below the experimental benchmark, suggesting that NSW participants may be negatively selected on some unobserved characteristic important for earnings.

Our re-analysis of the LaLonde (1986) dataset illustrates the use of AST in practice, we now report on a number of Monte Carlo experiments we conducted to verify the theoretical properties described in Theorem 3.2.1. In particular we wish to assess the relevance of our theoretical double robustness and efficiency results.

In each of our experiments we assume that $W$ is distributed according to a truncated normal distribution, with support $[-c, c]$, in both the study (treated) and auxiliary (control) populations. The location and scale parameters of these two distributions, respectively $(\mu_s, \sigma_s^2)$ and $(\mu_a, \sigma_a^2)$, may differ. We assume a multinomial sampling scheme: with probability $Q_0 = 1/2$ a draw of $(Y, W)$ is taken at random from the study population, otherwise a draw of $(X, W)$ is taken from the auxiliary population. Finally we assume that $Y$ and $X$, which play the roles of the outcome under treatment and control, are generated according to

$$\begin{aligned} Y|\, W, D &\sim \mathcal{N}\left(0, \sigma_Y^2\right) \\ X|\, W, D &\sim \mathcal{N}\left(\alpha_0 + \alpha_1\left(W - \mu_{W|D=1}\right) + \alpha_2\left[\left(W - \mu_{W|D=1}\right)^2 - \sigma_{W|D=1}^2\right], \sigma_X^2\right), \end{aligned}$$

where $\mu_{W|D=1}$ and $\sigma_{W|D=1}^2$ are the study population mean and variance of $W$ (which differ from $\mu_s$ and $\sigma_s^2$ due to truncation).

The target parameter is $\gamma_0 = \mathbb{E}_s[Y - X] = \alpha_0$. We vary $\sigma_Y^2$ to keep Hahn's (1998) semiparametric efficiency bound for $\gamma_0$ constant across designs (thus each design is asymptotically equivalent). The propensity score induced by these designs is of the logit form with an index quadratic in $W$ :

$$p_0(w) = \left[1 + \exp\left(-\beta_0 - \beta_1 W - \beta_2 W^2\right)\right]^{-1},$$

where $\beta_0$, $\beta_1$ and $\beta_2$ are functions of $(\mu_s, \sigma_s^2)$ and $(\mu_a, \sigma_a^2)$ (cf., Anderson, 1982). When the study and auxiliary population distributions of $W$ have different means, but a common variance, the logit index will be linear in $W$ (i.e., $\beta_1 \neq 0$ but $\beta_2 = 0$). When both the means and variances differ, then the index will generally be nontrivially quadratic in $W$ (i.e., $\beta_1 \neq 0$ and $\beta_2 \neq 0$).

Table 3.3 gives the parameter configurations for each of four Monte Carlo designs. In the first design both the propensity score, $p_0(w)$, and $q_a(w; \gamma_0)$ are 'linear' in $W$ (for $p_0(w)$ 'linear' means linear in the index). In the second design the propensity score is quadratic in $w$, while $q_a(w; \gamma_0)$ remains linear. In Design three the reverse is true, while in Design four both objects are 'quadratic'. Across each design we implement the AST estimator with $G(\cdot)$ being the logit function and $h(W) = W$. For parametric imputation (PI) we proceed 'as if' $\mathbb{E}[X|\, W]$ were linear in $W$, while our implementation of IPW uses a logit propensity score with a linear index, estimated by pseudo-maximum likelihood.

Our AST estimator is consistent for $\gamma_0$ in designs 1 through 3. PI is consistent in designs 1 and 2. IPW is consistent in designs 1 and 3, but inconsistent in design 2. All estimators are inconsistent in design 4. Table 3.4 reports the results of our experiments (with sample size set equal to 1,000 and 5,000 Monte Carlo replications). Column 1 lists a 'pencil and paper' asymptotic bias calculation, while Column 2 gives the median bias across the 5,000 Monte Carlo replications (in both cases bias is scaled by the 'pencil and paper' asymptotic standard error reported in Column 3). As predicted, AST is median unbiased (up to simulation error) across designs 1 through 3. In contrast, IPW is severely biased in design 2 and PI in design 3. As expected, all estimators perform poorly in design 4. These bias properties are reflected in the coverage of standard, Wald-based, 95 percent confidence intervals for $\gamma_0$ (Column 6).

The Monte Carlo experiments suggest three tentative conclusions: (i) the double robustness and efficiency results of Theorem 3.2.1 are evident in small samples, (ii) the lack of robustness in the PI and IPW estimators is consequential, and (iii) large sample approximations accurately characterize small sample performance (for the designs considered). Collectively we take these results to suggest the possibility of very real advantages of AST over other parametric procedures.

## 3.4 Discussion

Our AST estimator has antecedents in the statistics and econometrics literature. The interpretation of MD re-weighting under misspecification (i.e., when the imposed constraints are invalid) appears to have been first considered in the context of contingency table calibration by Little and Wu (1991). When misspecification is due to nonresponse in the sample being calibrated, they note a mapping between the calibration discrepancy and response probability for the case of logistic selection (p. 89). In the context of a creative proof of semiparametric just identification of their additive nonignorable (AN) attrition model, Hirano, Imbens, Ridder and Rubin (2001) fully generalize Little and Wu's (1991) observation (cf., Hellerstein and Imbens, 1999; Nevo, 2002). The present work provides an analogous mapping between calibration discrepancies and densities ratios appropriate for data combination problems.

Our work is also related to the application of nonparametric maximum likelihood methods to biased sampling problems with unknown biasing functions (e.g., Qin, 1998; Gilbert, Lele and Vardi, 1999). The solution to such problems, as in our work, requires imposing a prior restriction on a density ratio. Estimation in this literature proceeds by maximizing a nonparametric likelihood function with respect to a vector of multinomial probability weights and the finite-dimensional parameter indexing the biasing function (e.g., Gilbert, Lele and Vardi (1999), Equation 4.2, p. 32). Our approach to computation substantially differs. In our setup the parameters indexing the density ratio model correspond to Lagrange Multipliers in a *misspecified* minimum empirical discrepancy (MD) problem. An approach to estimation which begins with an intentionally misspecified MD problem is admittedly nonstandard. Nevertheless we provide conditions under which functionals of $F_s$ estimated efficiently. Furthermore, we provide conditions under which these estimates are consistent even if the (implicit) parametric form for the density ratio is misspecified. We view these efficiency and robustness features as our approach's primary valued-added.

The data combination problem, although related, differs from the class of semiparametric missing data problems studied by Robins, Rotnitzky and Zhao (1994). Hahn (1998) discusses the difference between the two problems in the context of program evaluation. Robins, Rotnitzky and Zhao (1994) propose an augmented inverse probability weighting (AIPW) estimator for such problems. The AIPW estimator is also locally efficient and 'doubly robust' (Tsiatis, 2006). The AIPW estimator is an M-estimator based on a parametric estimate of the efficient score for the missing data model; what Newey (1990) terms a 'quasi-efficient score'. Forming the quasi-efficient score requires a model for the probability of missingness (i.e., the propensity score) and the conditional expectation of the identifying moment given always-observed covariates. Double robustness of the Robins, Rotnitzky and Zhao (1994) estimator stems from the fact that mean-zeroness of the quasi-efficient score requires only one of these two auxiliary parametric models to be correctly specified.

The efficient score function for the data combination model considered here was derived by Hahn (1998) for the special case where $\gamma_0$ corresponds to ATT and for the general moment condition case by Chen, Hong and Tarozzi (2008). The analog of the AIPW estimator for data combination problems (i.e., a M-estimator based on a parametric estimate of the efficient score) does not have a double robustness property. This is related to Hahn's (1998) finding that smoothness priors on the propensity score change the efficiency bound for such problems. This suggests that double robustness of our MD procedure is somewhat surprising.

Table 3.1: Difference between means of variables amongst NSW participants and CPS 'controls' before and after reweighting

| | (1) NSW $D=1$ | (2) CPS $D=0$ | (3) (1)-(2) | (4) IPW (1)-(2) | (5) AST (1)-(2) |
|---|---|---|---|---|---|
| **Pre-Intervention Characteristics** | | | | | |
| Age | 25.98 | 32.66 | -6.68 | 1.34 | 0.00 |
| Years-of-Schooling | 10.31 | 10.98 | -0.67 | 0.05 | 0.00 |
| Married | 0.19 | 0.62 | -0.43 | 0.02 | 0.00 |
| Dropout | 0.72 | 0.43 | 0.29 | 0.03 | 0.00 |
| Zero Earnings in 1974 | 0.71 | 0.14 | 0.57 | -0.01 | 0.00 |
| Zero Earnings in 1975 | 0.61 | 0.16 | 0.46 | -0.00 | 0.00 |
| Zero Earnings in 1974 & 1975 | 0.60 | 0.10 | 0.50 | -0.00 | 0.00 |
| Earnings in 1974 | 2.16 | 11,43 | -9.27 | 0.18 | 0.00 |
| Earnings in 1975 | 1.49 | 10.94 | -9.45 | 0.10 | 0.00 |
| Earnings in 1974 squared | 30.62 | 208.93 | -178.30 | 3.41 | 0.00 |
| Earnings in 1975 squared | 13.02 | 194.55 | -180.53 | 0.73 | 0.00 |
| Product of 1974 and 1975 earnings | 14.37 | 192.99 | -178.62 | 1.39 | 0.00 |
| **Post-Intervention Earnings** | | | | | |
| Earnings in 1978 | 6.14 | 12.01 | - | - | - |
| N | 156 | 1,176 | | | |

NOTES: Earnings are expressed in thousands of 1982 dollars. 'Dropout' is a binary variable equal to one if the individual failed to complete high school and zero otherwise. Additional details regarding the data can be found in LaLonde (1986). The data were downloaded from http://www.nber.org/˜rdehejia/nswdata.html in May of 2009.

Table 3.2: Observational Estimates of the ATT: NSW Demonstration

| (1) | (2) | (3) | (4) |
|---|---|---|---|
| **Exp** | **PI** | **IPW** | **AST** |
| 2,029 | 1,657 | 1,302 | 1,284 |
| (748) | (755) | (787) | (785) |

NOTES: The 'Experimental' ATT estimate is a difference between the raw means of the 1978 earnings of NSW participants and non-participants in the NSW evaluation dataset. The included group of participants ($N_s$= 156) and non-participants ($N_a$= 371) are those who are (i) black and (ii) for whom earnings information in both 1974 and 1975 is available. Standard errors are reported in parentheses. The remaining estimates are based on the merged sample of NSW participants and CPS controls.

Table 3.3: Parameter values for the four Monte Carlo experiments.

| Design | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $\mu_s$ | 0 | 0 | 0 | 0 |
| $\sigma_s^2$ | 1 | 1 | 1 | 1 |
| $\mu_a$ | -1/2 | -1/2 | -1/2 | -1/2 |
| $\sigma_a^2$ | 1 | $\sqrt{3/4}$ | 1 | $\sqrt{3/4}$ |
| $c$ | 3 | 3 | 3 | 3 |
| $\sigma_Y^2$ | 3.4823 | 3.0928 | 1.7496 | 1.3601 |
| $\alpha_0$ | 0 | 0 | 0 | 0 |
| $\alpha_1$ | 1/2 | 1/2 | 1/2 | 1/2 |
| $\alpha_2$ | 0 | 0 | -1 | -1 |
| $\sigma_X^2$ | 1 | 1 | 1 | 1 |
| $\sqrt{\mathcal{I}(\gamma_0)^{-1}/N}$ | 0.1 | 0.1 | 0.1 | 0.1 |

NOTES: The square root of Hahn's (1998) variance bound for each design (divided by $N^{1/2} = \sqrt{1,000}$) is reported in the last row of the table.

Table 3.4: Monte Carlo results

| | (1) Asymptotic Bias | (2) Median Bias | (3) Asymptotic Std. Err. | (4) Median Std. Err. | (5) Standard Deviation | (6) Coverage of 95% CI |
|---|---|---|---|---|---|---|
| | **Design 1: $p_0(w)$ linear, $q_a(w;\gamma_0)$ linear** | | | | | |
| **PI** | 0.0000 | 0.0097 | 0.0997 | 0.0996 | 0.0986 | 0.9526 |
| **IPW** | 0.0000 | 0.0164 | 0.1007 | 0.1006 | 0.1005 | 0.9506 |
| **AST** | 0.0000 | 0.0057 | 0.1000 | 0.0998 | 0.0998 | 0.9542 |
| | **Design 2: $p_0(w)$ quadratic, $q_a(w;\gamma_0)$ linear** | | | | | |
| **PI** | 0.0000 | 0.0277 | 0.0966 | 0.0966 | 0.0962 | 0.9492 |
| **IPW** | 0.3398 | 0.3780 | 0.0958 | 0.0960 | 0.0958 | 0.9334 |
| **AST** | 0.0000 | 0.0302 | 0.0972 | 0.0970 | 0.0969 | 0.9492 |
| | **Design 3: $p_0(w)$ linear, $q_a(w;\gamma_0)$ quadratic** | | | | | |
| **PI** | -1.6125 | -1.6308 | 0.1309 | 0.1296 | 0.1321 | 0.6204 |
| **IPW** | 0.0000 | -0.0137 | 0.1063 | 0.1037 | 0.1068 | 0.9420 |
| **AST** | 0.0000 | -0.0380 | 0.1125 | 0.1105 | 0.1130 | 0.9444 |
| | **Design 4: $p_0(w)$ quadratic, $q_a(w;\gamma_0)$ quadratic** | | | | | |
| **PI** | -3.7266 | -3.7566 | 0.1242 | 0.1214 | 0.1238 | 0.0288 |
| **IPW** | -2.0255 | -2.0702 | 0.0927 | 0.0902 | 0.0932 | 0.4492 |
| **AST** | -2.1528 | -2.1942 | 0.1061 | 0.1002 | 0.1046 | 0.3846 |

NOTES: Each row corresponds to a specific estimator as described in the main text. Column 1 reports the scaled large sample bias of each estimator (i.e., its probability limit minus the true parameter divided by the square root of its large sample variance, $(AVar(\widehat{\gamma})/N)^{1/2}$). Column 2 reports the median Monte Carlo bias of each estimator scaled by its asymptotic standard error. Column 1 calibrates the scale of inconsistency for each estimator, while a comparison of Columns 1 and 2 allows for an assessment of whether an estimator's actual sampling distribution is centered at is probability limit. Column 3 gives the large sample standard error of each estimator (i.e., $(AVar(\widehat{\gamma})/N)^{1/2}$), Column 4 the median estimated standard error and column 5 the standard deviation of the point estimates across the 5,000 Monte Carlo replications. Column 6 reports the actual coverage of a 95 percent Wald-based confidence interval.

# Bibliography


Alagbari, B. R. (1992). نظام التعليم في الجمهورية اليمنية *(The Educational System in the Republic of Yemen)*. صنعاء: دار اقراء.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature 43*(3), 762–800.

Alwashli, Y. (2007). دراسة في اليمن: سياسية بناء قوة الدولة، دراسة جيوستراتيجية *(A Study in Yemen: Builiding the Political Power of the State, A Geostrategic Study)*. صنعاء.

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* (18), 47–82.

ARD (2006). Yemen Corruption Assessment. Technical report, US AID.

Arellano, M. and S. Bond (1991). Some specification tests for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–298.

Banerjee, A., D. Mookherjee, K. Munshi, and D. Ray (2001). Inequality, control rights, and rent seeking: Sugar cooperatives in maharashtra. *Journal of Political Economy 109*(1).

Bardhan, P. (1997). Corruption and development: a review of issues. *Journal of Economic Literature 35*(3), 1320–1346.

Bardhan, P. and D. Mookherjee (2005). *Handbook of Economic Corruption*, Chapter Decentralization, Corruption And Government Accountability: An Overview.

Black, S. E. and K. L. Sokoloff (2006). Long-term Trends in Schooling: The Rise and Decline (?) of Public Education in the United States. In E. A. Hanusek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1, Chapter 2.

Boase, B. (2001). The path toward local administration for the republic of yemen. Technical report, The Government of the Republic of Yemen and the Donor Community.

Boli, J., F. O. Ramirez, and J. W. Meyer (1985). Explaining the origins and expansion of mass education. *Comparative Education Review 29*(2), 145–170.

Braggion, F. (2009, March). Social networks eects on listed and unlisted firms: Evidence from managers aliations with freemasonry. Mimeo.

Burchfield, M., H. G. Overman, D. Puga, and M. A. Turner (2006). Causes of sprawl: A portrait from space. *Quarterly Journal of Economics 121*(2), 587–633.

Burt, R. (2003). Freemasonry and business networking during the victorian period. *Economic History Review LVI*(4), 657–688.

Chandler, A. (1977). *The Visible Hand*. Cambridge, Mass. and London, England: The Belknap Press of Harvard University Press.

Chaney, E. (2008, November). Tolerance, religious competition and the rise and fall of muslim science.

Charrad, M. (2001). *States and Women's Rights: The Making of Postcolonial Tunisia, Algeria, and Morocco*. Berkeley: University of California Press.

Clark, G. B. (1936). *Our Masonic Heritage: A Glimpse of the Historical Background of Freemasonry*. Denver: Grand Lodge A.F. and A.M. of Colorado.

Cohen, J. M., M. Hébert, D. B. Lewis, and J. C. Swanson (1981). Development from below: Local development associations in the yemen arab republic. *World Development 9*(11/12), 1039–1061.

Contin, R., D. Egel, K. Moore, and H. Ogleh (2009). Chf yemen access-plus baseline report: A report on child labor and education in aden, hajja, hudeida and taiz governorates in yemen. Technical report, CHF International.

CPO, C. (1974). *Statistical Yearbook*. Sanaa.

Crook, R. C. (2003). Decentralisation and poverty reduction in africa: The politics of local-central relations. *Public Administration and Development 23*, 77–88.

Cubberley, E. (1920). *The History of Education*. Houghton Mifflin Company.

Cubberley, E. (1922). *A Brief History of Education: A History of the Practice and Progress and Organization of Education*. Houghton Mifflin Company.

Daair, O. (2001). He who rides the lion. authortitarian rule in a plural society: The republic of yemen. Master's thesis, University of London, http://www.al-bab.com/Yemen/pol/daair1.htm.

Davis, J. S. (1917). *Essays in the Earlier History of American Corporations*. Harvard University Press.

Day, S. (2001). *Power-Sharing and Hegemony: A Case Study of the United Republic of Yemen*. Ph. D. thesis, Georgetown University.

De Los Reyes, G. and A. Lara (1999). Civil society and volunteerism. *ANNALS, AAPSS 565*.

Demott, B. J. (1986). *Freemasonry in American Culture and Society*. University Press of AMerica.

Dexter, E. (1904). *A History of Education in the United States*. The Macmillan company.

Dresch, P. (1986). Significance of the course events take in segmentary systems. *American Ethnologist 13*(2), 309–324.

Dresch, P. (1989). *Tribes, Government and History in Yemen*. Oxford University Press.

Dresch, P. (1995). The tribal factor in the yemeni crisis. In *The Yemeni War of 1994: Causes and Consequences*, pp. 33–55. Emirates Centre for Strategic Studies and Research: Saqi Books.

DRI (2008, November 2008). Yemen: Comprehensive assessment of eleciton framework. Final report, Democracy Reporting International.

Dumenil, L. (1984). *Freemasonry and American Culture 1880-1930*. Princeton University Press.

Dumenil, L. (1990). "The Insatiable Maw of Bureaucracy": Antistatism and Education Reform in the 1920s. *The Journal of American History 77*(2), 499–524.

Easterlin, R. A. (1960). Interregional differences in per capita income, population, and total income, 1840-1950. In *Conference on Research in Income and Wealth, Trends in the American Economy in the Nineteenth Century, Studies in Income and Wealth*, Volume 24, pp. 73–140. Princeton University Press.

Eby, F. (1925). *The Development of Education in Texas*. New York: The MacMillan Company.

Evans-Pritchard, E. (1940). *The Nuer: A Description of the Modes of Livelihood and Political Institutions of a Nilotic People*. Oxford, England: Oxford University Press.

Fels, T. (2002). Religious assimilation in a fraternal organization. In R. Weisberger, W. McLeod, and S. Morris (Eds.), *Freemasonry on Both Sides of the Atlantic: Essays Concering the Craft in the British Isles, Europe, the United States and Mexico*. Columbia University Press.

Fox, W. L. (1997). *Lodge of the Double-Headed Eagle: Two Centuries of Scottish Rite Freemasonry in America's Southern Jurisdiction*. University of Arkansas Press.

Gellner, E. (1981). *Muslim Society*. Cambridge University Press.

Gellner, E. (1991). *Tribes and State Formation in the Middle East*, Chapter Tribalism and the state in the Middle East, pp. 109–126. Tauris.

Go, S. (2008). Free schools in america, 1850-1870: Who voted for them, who got them, and who paid. *Job Market Paper, University of California, Davis*.

Go, S. and P. H. Lindert (2008, November 10). The Curious Dawn of American Public Schools. University of California - Davis.

Goldin, C. and L. F. Katz (1998, May). The origins of state-level differences in the public provision of higher education: 1890-1940. *The American Economic Review*, 303–308.

Goldin, C. and L. F. Katz (1999). Human capital and social capital: The rise of secondary schooling in america, 1910-1940. *Journal of Interdisciplinary History XXIX*(4), 683–723.

Goldin, C. and L. F. Katz (2003). Mass secondary schooling and the state: The role of state compulsion in the high school movement. NBER Working Paper # 10075.

Goldin, C. and L. F. Katz (2008). Why the United States Led in Education: Lessons from Secondary School Expansion, 1910 to 1940.

Goldsmith, R. W. (1952). The Growth of Reproducible Wealth of the United States of America from 1805 to 1950. *Review of Income and Wealth 2*(1), 247–328.

Greif, A. (1993). Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition. *American Economic Review 83*(3), 525–48.

Harland-Jacobs, J. (2007). *Builders of Empire: Freemasons and British Imperialism, 1717-1927*. UNC Press.

Haywood, H. L. (1963). *The Masonic Essays of H. R. Haywood*, Volume 20. Fulton: Transactions of the Missouri Lodge of Research.

Hope, K. G. (2005). "One Can Make A Difference". *The California Encompasser 19*(4).

IFES (2005). Election law reform in yemen: Final report on the public policy dialogues. Technical report, IFES-Yemen.

Khoury, P. and J. Kostiner (Eds.) (1990). *Tribes and State Formation in the Middle East*. London: Tauris.

Kodi, M. W. (2007). Anti-corruption challenges in post-election democratic republic of congo. Technical report, Chatham House.

Kuran, T. (2004). Why the middle east is economically underdeveloped: Historical mechanisms of institutional stagnation. *The Journal of Economic Perspectives 18*(3), 71–90.

Kutolowski, K. S. (1982). Freemasonry and community in the early republic: The case for antimasonic anxieties. *American Quarterly 34*, 543–561.

Kuznets, S. (1946). *National Product Since 1869*. (Assisted by Lillian Epstien and Elizabeth Jenks). National Bureau of Economic Research.

Lang, O. (2006). *History of Freemasonry in the State of New Yor*. Kessinger Publishing.

Lindert, P. H. (2004). *Growing public : social spending and economic growth since the eighteenth century*. Cambridge University Press.

Mackey, A., R. Clegg, and H. Haywood (1946). *Encyclopedia of Freemasonry*. Kessinger Publishing.

Mackey, A. and H. Haywood (2003). *Encyclopedia of Freemasonry Part 2*. Kessinger Publishing.

Mauro, P. (1995). Corruption and growth. *Quarterly Journal of Economics 110*(3), 681–712.

Meissner, J. (1987). *Tribes at the Core: Legitimacy, Structure and Power in Zaydi Yemen*. Ph. D. thesis, Columbia University.

Meon, P.-G. and K. Sekkat (2005). Does corruption grease or sand the wheels of growth? *Public Choice 122*, 69–97.

Moore, C. (1846). *The Freemasons' Monthly Magazine*.

Moulton, B. R. (1990, May). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics 72*(2), 334–38.

Mundy, M. (1995). *Domestic Government: Kinship, Community and Polity in North Yemen*. London: Tauris.

NDI (2000). Statement of the NDI Pre-election Delegation to Yemen's April 2001 Parliamentary Election. Technical report, National Democratic Institute for International Affairs, Sana'a.

NDI (2007). Yemen: Tribal conflict management program research report. Technical report, National Democratic Institute for International Affairs.

Nugent, J. B. and N. Sanchez (1999). The local variability of rainfall and tribal institutions: The case of sudan. *Journal of Economic Behavior & Organization 39*, 263–291.

Nunn, N. and D. Puga (2009). Ruggedness: The Blessing of Bad Geography in Africa. *NBER Working Paper 1498*.

Odhiambo, M., W. Mittulah, and K. Akivaga (2005). *Management of resources by local authorities: the case of the Local Authority Transfer Fund in Kenya*. Nairobi: Clarion Press.

Parry, D. L. L. (1998). Friends in high places: The favours sought by the freemasons of orleans, 1890-1914. *French History 12*(2).

Phillips, S. (2008). *Yemen's Democracy Experiment in Regional Perspective*. Palgrave Macmillan.

Pike, A. (1871). *Morals and dogma of the Ancient and Accepted Scottish Rite of Freemasonry*. Masonic Publishing Company.

Posusney, M. P. and M. P. Angrist (Eds.) (2005). *Authoritarianism in the Middle East: regimes and resistance*. Boulder, Colorado: Lynne Rienner Publishers.

Poterba, J. M. (1997). Demographic structure and the political economy of public education. *Journal of Policy Analysis and Management, 16*(1), 48–66.

Reinikka, R. and J. Svennson (2006). Using Micro-Surveys to Measure and Explain Corruption. *World Development 34*(2), 359–370.

Riley, S. J., S. D. DeGloria, and R. Elliot (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences 5*(1-4), 23–27.

Robinson, J. A. and T. Verdier (2003). The political economy of clientelism.

Romeo, L. G. and M. El Mensi (2008, September). The difficult road to local autonomy in yemen.

Rosenzweig, R. (1977). Boston masons, 1900-1935: the lower middle class in a divided society. *Nonprofit and Voluntary Sector Quarterly 6*(119).

Royal Arch Masons of California (1882). *Proceedings of the Grand Chapter of Royal Arch Masons of the State of California at Its Annual Convocation.* Royal Arch Masons of California: Royal Arch Masons, Grand Chapter.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010). Integrated public use microdata series: Version 5.0 [machine-readable database]. Technical report, Minneapolis: University of Minnesota.

Sappington, J. M., K. M. Longshore, and D. B. Thompson (2007). Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study Using Bighorn Sheep in the Mojave Desert. *Journal of Wildlife Management 71*(5), 1419–1426.

Seaman, E. C. (1868). *Essays on the Progress of Nations, in Civilization, Productive Industry, Wealth and Population: Illustrated by Statistics of Mining, Agriculture, Manufactures, Commerce, Coin, Banking, Internal Improvements, Emigration, and Population.* C. Scriber & Co.

Shleifer, A. and R. Vishny (1993). Corruption. *Quarterly Journal of Economics 108*(3), 599–617.

Soltow, L. and E. Stevens (1981). *The Rise of Literacy and the Common School in the United States.* University of Chicago Press, Chicago.

Spinelli, A. (2003). The voter registration system in yemen. Technical report, UNDP Electoral Support Project.

Stevenson, T. B. (1985). *Social Change in a Yemeni Highlands Town.* University of Utah Press.

Stillson, H. L. and W. J. Hughan (1890). *History of the Ancient and Honorable Fraternity of Free and Accepted Masons, and Concordant Orders.* The Fraternity publishing company.

Stookey, R. W. (1974, Summer). Social Structure and Politics in the Yemen Arab Republic - Part I. *Middle East Journal 28*(3), 248–260.

Swagman, C. (1988). *Development and Change in Highland Yemen.* Salt Lake City: University of Utah Press.

Tabbert, M. A. (2005). *American Freemasons: Three Centuries of Building Communities.* National Heritage Museum and New York University Press.

Tapper, R. (1990). Anthropologists, Historians and Tribespeople on Tribe and State Formation in the Middle East. In P. Khoury and J. Kostiner (Eds.), *Tribes and State Formation in the Middle East*. Berkeley and Los Angelese, California: University of California Press.

Thornton, J. (1962). The Contributions of Freemasonry in Texas to Public School Education, 1846 to 1861. Master's thesis, Baylor University, Waco.

Thornton, J. (1970). *The Relationship of Freemasonry to the Structure, Organization, and Administration of Public School Education in Texas, 1861-1885*. Texas A & M University.

Tucker, G. (1843). *Progress of the United States in Population and Wealth in Fifty Years, as Exhibited by the Decennial Census: As Exhibited by the Decennial Census*. Little & Brown.

Véron, R., G. Williams, S. Corbridge, and M. Srivastava (2006). Decentralized corruption or corrupt decentralization? community monitoring of poverty-alleviation schemes in eastern india. *World Development 34*(11), 1922–1941.

Webb, W. (1955). *Masonry in Texas, Background, History and Influence in 1846*. Fort Worth: The Masonic Home and School Press.

Weber, M. (1946). *The Protestant Sects and the Spirit of Capitalism*. New York: Oxford University Press.

Weir, S. (2007). *A Tribal Order: Politics and Law in the Mountains of Yemen*. University of Texas Press.

Whitsell, L. O. (1950). *One Hundred Years of Freemasonry in California*. Grand Lodge F. & A. M. of California.

Woods, R. O. (1936). Masonic educational institutions. *The School Review 44*(6), 456–461.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

World Bank (2006). Tracking basic education expenditures in yemen: Analyses of public resources management and teacher absenteeism. Technical report, Social and Economic Development Group, Middle East and North Africa Region.

# Appendix A

# Appendices to Chapter 1

## A.1 The Recent Expansion of Yemen's Education System

The education system that exists today in the north of the Republic of Yemen is a very recent phenomenon. The first public schools did not open there doors until the early 1960s, and while the educational system did expand rapidly from the late 1970s until the present, the overall penetration of the education system into rural areas is still quite limited. Importantly, given its recent arrival and the stability and longevity of the tribes discussed above, it is unlikely that the education system has affected the structure of the tribe.

In northern Yemen in 1962, the year of the revolution which saw the removal of a religious monarchy, there were only 23 schools serving a population of over 4 million.[1] And while the education system did expand in the wake of the 1962 revolution, the real expansion of the education system did not happen until the late 1970s.[2] The delayed arrival of widespread access to education can be seen dramatically in the extremely high illiteracy rates among adults, shown in Figures 1.1 (men), that exceed 60% in rural areas and are among the very highest in the world.[3]

The rapid expansion of the education system in the late 1970s, and its continued expansion through the 1980s and 1990s, is illustrated in Figures 1.3 and 1.10 which show, respectively, the construction of schools and classrooms during the 20th century.[4] Today, with over 16,000 schools in the education system, there is a school for every third village. However, despite the apparent widespread availability of schools, and the significant resources that are spent on education[5], enrollment rates are still quite low with 68% of eligible boys and 45% of eligible girls currently enrolled.[6] And, importantly, the relatively dramatic geographic

[1]Under the rule of the Imam, education was done almost entirely in mosques, focused on religious study and the study of Islamic jurisprudence and was restricted to only the elite.

[2]Though Arab and other foreign countries did play an important role in the provision of teachers and the development of curriculum in the early years of this expansion, as is often discussed, local communities also played a central role (Alagbari 1992). Indeed, nearly 60% of the school capacity built in Yemen was built by local communities without the financial support of either the central government or foreign donors (author's calculations using the 2007 Education Census)

[3]The standard census definition of an adult in Yemen is defined as any individual over 15.

[4]The initial expansion of the education system was driven by the first oil boom as the remittances earned by Yemeni workers employed in Saudi Arabia and the Gulf surged. Alwashli (2007) provides a more expansive discussion of the expansion of the education system after the Revolution.

[5]Education is the largest component of current government expenditures and accounts for around 20% of the total budget (though the actual share varies substantially from year to year though it has remained between 15-25% of the budget in recent years).

[6]In addition to lack of access to educational facilities, two central explanations are usually

variation in the enrollment rates in rural areas, as shown in Figures 1.2, demonstrates the variance in educational opportunities available to Yemeni children in these areas.

The recent expansion of the education system, which has occurred almost entirely in the last 20-25 years as demonstrated in Figures 1.3 and 1.10, is key to my identification assumption. Indeed, while the education system, and the patronage system that accompanies it as discussed in Section 1.5.1 below, has developed very recently, the tribal system is both old and stable as discussed in Section 1.2.3 above. It is thus unlikely that the education system would have impacted the basic tribal structure in any way.

## A.2 Data Appendix

### A.2.1 Yemen's Administrative Structure

The Republic of Yemen currently has three main levels in its administrative hierarchy.[7] The largest administrative structure is the governorate of which there are a total of 21, six of these are from the former People's Democratic Republic of Yemen and the remaining 15 are from the former Yemen Arab Republic (North Yemen). These 21 governorates are then sub-divided into a total of 333 districts which are further divided into nearly 2,200 sub-districts, the smallest official administrative structure. The nearly 40,000 villages, which are themselves composed of approximately 200,000, do not have any official status and are typically a locally defined concept.[8]

For the rural populations that are the focus of this study, the most recent population census from 2004 reported a total population of just over 14 million

offered for these low enrollment rates. The first is the high cost of education. Though school is purportedly freely provided, a variety of fees are typically charged to students which are often prohibitive (see Contin, Egel, Moore, and Ogleh (2009) for a discussion of this). Note that the Ministry of Education in cooperation with the World Bank and the European Council are currently experimenting with several conditional cash transfer programs to help alleviate this difficulty. The second is the particular severity of teacher absenteeism in Yemen with estimates of absenteeism ranging from 16% (World Bank 2006) to around 50% (Contin, Egel, Moore, and Ogleh 2009).

[7]A variety of other structures, such as sub-governorates, were used in the past but do not find much practical use today.

[8]It is thus impossible to calculate the actual number of villages in Yemen as villages are defined differently in different data sources. Indeed, while a particular hamlet may be reported in a census as part of another village by a local informant, the informant that is interviewed in another census may indicate that the local hamlet is actually another village. In general, the census officers and field workers defer to the judgement of the local informant which is typically a local elder or leader. In many cases this leads to settlements with only one household and 8-10 members being identified as a separate village in the census.

individuals residing in 38,736 villages.[9] The median district had a population of approximately 35,000, though there is significant variation in the population of these governorates as the smallest governorate had a population of under 2,000 while the largest had a population of nearly 200,000. And the median sub-district had 4,000 inhabitants though again there was relatively significant variation of over 8000 with sub-district populations ranging from only one hundred individuals to approximately 80,000.[10]

### A.2.2 Population and Economic Controls

The population censuses from 1994 and 2004, which were used for the calculation of the true enrollment rate within an area, contain a variety of other important population and economic control variables. They report the size of the potential local student population as well as a variety of variables designed to measure the amount of public services available in these areas. The public services surveyed include (1) percent of citizens without access to sanitation, (2) percent of citizens without access to electricity, (3) percentage of households using wood, coal and kerosene for cooking, (4) percentage of households without water from a paper and (5) the educational enrollment rates of boys and girls ages 6-15. While these data are available at the village level, sub-district average/aggregates were calculated for the analysis considered here using the total population in a village as weights. The summary statistics for these variables are included in Table 1.4.

### A.2.3 Agricultural Controls

Agriculture, which is the primary source of income in rural Yemen with the exception of migrant labor, undoubtedly had an impact on the development of the tribal structure. And as agricultural production continues to be an important source of local wealth, and an important cause of local inequality, it might be expected to affect modern economic and political outcomes. Access to cultivable land, the availability of water resources, the type of land sharing arrangements and the type of crops that can be supported by the local climate would all be expected to affect individual returns to education as well as the overall development of the community.

In order to control for the potential impact of agricultural factors on both tribes and educational variables, I include a variety of agricultural variables, drawn from the 2001 Agricultural Census. These variables include: the total amount of land owned by private individuals, the share of the land that is cultivable, the

[9]The total population in 2004 was just over 20 million.

[10]In the southern govenorates, i.e. those of former South Yemen, sub-districts are much more rare and the district is often the smallest administrative structure above the village.

type of water access that is available, the size of land holdings, the amount of animal assets held and the amount of land devoted to grain, qat and cash crop production. This last variable is of particular important as *qat*, which requires significant amounts of water and grows only in specific climates, is the only true cash crop in Yemen. Though these data are available at the village level, as the primary unit of analysis here is the sub-district, I calculate sub-district averages for each of the variables using the total population in each village as the weight for that village. The full list of variables as well as their mean and standard deviation are included in Table 1.4.

The agricultural census does not include all of the governorates for which I have tribal data. The governorate of Al-Jawf was not included in the agricultural census as the Ministry of Agriculture and Central Statistical Office judged that the data from this governorate were systematically biased (i.e. residents of these areas were instructed to lie about their assets and land by their governor). This is a particular concern as the villages code that are used to match the 1994 population census are drawn from the agriculture census, and I am thus compelled to drop this governorate throughout my analysis.

### A.2.4 Area and Terrain Controls

Both population density and terrain quality are likely to affect both educational outcomes and the tribal structure. Thus, my analysis includes measures of terrain ruggedness as well as the the area of the units of analysis. The latter variable corrects for the population density as all the analysis includes measures of the size of the local population. See Table 1.4 for the summary statistics of these variables.

Measures of area were drawn from two sources. The area of the administrative units of analysis themselves which was extracted from the ArcGIS maps that accompanied the 2004 population census, and the 2001 agriculture census reports the total amount of land claimed by the villages contained within an administrative unit. In all cases, the logarithm of area was calculated before it was included in the analysis.

For my measure of terrain ruggedness I use the vector ruggedness measure (VRM) of Sappington, Longshore, and Thompson (2007).[11] Though other studies in economics have focused on the terrain ruggedness index (TRI) of Riley, DeGloria, and Elliot (1999) (cf. Burchfield, Overman, Puga, and Turner 2006; Nunn and Puga 2009), there are two reasons that I have opted to use the VRM. The

[11] The USGS GTOPO30 file 'e020n40', which reports the elevation at approximately one kilometer intervals for the Arabian peninsula and eastern Africa, was used for these calculations. It was downloaded from `http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info`.

first is that the VRM is more appropriate for the current analysis as it quantifies ruggedness independently of slope. This is important because even steep terrain is relatively easy to traverse if it is not uneven and broken. Second, calculation of the VRM is facilitated by the availability of a publicly available toolbox for ArcGIS that is designed to calculate this measure.[12]

## A.3 Elite Capture during the Decentralization Process

Three aspects of the decentralization process gave local elites significant control over the results of this process. A first major concern was that the demarcation of the local council constituencies was particularly susceptible to manipulation by local, typically tribal, elites. Though the Local Authority Law did create relatively specific criteria for the size of the constituencies for the local elections, in that all constituencies within an administrative border must be within 5% of each other in terms of population, the Supreme Elections Committee (SEC) had neither sufficient time nor the capacity to properly delineate these constituencies. Indeed, as the law authorizing the elections was only passed in 2000, and the elections were held the following year, the SEC was left with the responsibility of creating nearly 7,000 new electoral districts within a year (NDI 2000). A daunting task without the added complication that the SEC did not (1) have the demographic data easily available for constructing these districts and (2) did not have the technical capability to use any sort of technology to help them in the process. The result was that the SEC became overly reliant on the the assistance of local, and often tribal, elites in creating these constituencies.

Second, there is an indication that voter registration and education for these elections was insufficient to guarantee democratic outcomes, and that their weakness gave local elites significant control over the outcome of elections. First, the voter registration system had a number of important deficiencies. These deficiencies included the disqualification of voters who were unregistered as of 1999, the existence of duplicate and false registrations,[13] and the ability of individuals to vote where they lived, worked or were born. More nefarious irregularities including voter intimidation and voter impersonation, which was relatively common,

[12] This toolbox is available from `http://arcscripts.esri.com/details.asp?dbid=15423`. It is important to note that the TRI measure also has a script available to facilitate calculation (`http://arcscripts.esri.com/details.asp?dbid=12435`). However, this script is in practice quite difficult to implement with ArcGIS desktop as it was developed for ArcInfo Workstation.

[13] The report from the National Democratic Institute (2000) noted the existence of both dead and fictitious people on voting rosters. And while nearly 200,000 of these names were cleansed from official registries before the local elections, it is certain that many more remained.

also occurred during these elections.[14] Second, the effectiveness of these local elections in selecting councils that are representative of the local populations is predicated on the existence of local populations that are both educated about the elections and their implications. However, as the local elections were announced only a year before they took place, observers of the election were concerned that the SEC was largely incapable of meeting its legal responsibility to educate the public about the governorate and local council elections (NDI 2000).

A third concern is that the nomination process for candidates of local elections highly favored the existing elite. While, in principle, all local residents were qualified to contest elections for the local councils, two important factors served to restrict the candidate pool. The first is that local election committees would often reject nomination applications without explanation and without any possibility for that individual to appeal (IFES 2005). This issue is particularly significant as local election committees were appointed and typically from the local elite. The second is that individuals who were not supported by one of the major political parties faced difficulty in the nomination process as they were required to obtain signatures that needed to be validated by a member of the local elite (i.e. local elder, local judge) before their nomination would be accepted.

[14]The IFES (2005) report alludes to the coercion of voters in their recommendations for improvements in the voting system. Spinelli (2003) refers to the existence of significant amounts of voter impersonation in his recommendations.

# Appendix B

# Appendices to Chapter 2

## B.1 Freemason Data

| | State Founded | Economic Data First Available | First Lodge† | Grand Lodge Established | Years for Which Freemason Data is Available | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 1850 | 1860 | 1870 | 1880 | 1890 | 1900 |
| Arizona | 1912 | 1870 | 1866 | 1882 | | | X | X | X | X |
| Arkansas | 1836 | 1850 | 1835 | 1837 | X | X | X | X | X | ?? |
| California | 1850 | 1850 | 1850 | 1850 | X | X | X | X | X | X |
| Colorado | 1876 | 1860[1] | 1860 | 1861 | | X | X | X | X | X |
| Idaho | 1890 | 1870 | 1863 | 1867 | | | X | X | ?? | X |
| Illiois | 1818 | 1850 | 1838 | 1840 | X | X | X | X | X | X |
| Indiana | 1816 | 1850 | 1807 | 1818 | X | X | X | X | X | X |
| Iowa | 1846 | 1850 | 1840 | 1841 | X | X | X | X | X | X |
| Kansas | 1861 | 1860 | 1854 | 1856 | | X | X | X | X | X |
| Louisiana | 1812 | 1850 | 1794 | 1812 | ?? | X | X | X | X | X |
| Minnesota | 1858 | 1850 | 1852 | 1853 | | X | X | X | X | X |
| Missouri | 1821 | 1850 | 1808 | 1821 | X | ?? | X | X | X | X |
| Montanta | 1889 | 1870 | 1864 | 1866 | | | X | X | X | X |
| Nebraska | 1867 | 1860 | 1855 | 1857 | | X | X | X | X | X |
| Nevada | 1864 | 1860 | 1864 | 1866 | | | X | X | X | X |
| New Mexico | 1912 | 1850 | 1851 | 1879 | | ** | X | X | X | X |
| North Dakota | 1889 | 1870 | 1874[2] | 1890 | | | | X | X | X |
| Oregon | 1859 | 1850 | 1848 | 1851 | X | X | X | X | X | X |
| South Dakota | 1889 | 1860 | 1863 | 1875 | | | X | X | X | X |
| Texas | 1845 | 1850 | 1836 | 1838 | ?? | X | X | X | ‡ | ?? |
| Utah | 1896 | 1850 | 1866[3] | 1872 | | | X | X | X | X |
| Washington | 1889 | 1860 | 1852 | 1858 | X | X | X | X | X | X |
| Wisconsin | 1848 | 1850 | 1843 | 1843 | X | X | X | X | X | X |
| Wyoming | 1890 | 1870 | 1868 | 1874 | | | X | X | X | X |

Notes:
??: Data is in principle available but was not available at the Iowa Masonic Library.
**: Only one lodge was open in New Mexico in 1860 (under the jurisdiction of the Grand Lodge of Missouri). The data for this year was estimated as half that of 1870. As the 1860 Freemason is not used for quantitative analysis, this assumption does not affect any of the results (just the map).
‡: Data for Texas was not available for 1890. 1888 was the last year for which Freemason membership data was available.
†: There was often a difference of a year or more between the data in which a lodge was established and which it was chartered.
[1]: There is only very limited data available for Colorado in 1860 - only one county with data.
[2]: There was one lodge formed slightly earlier, but it was closed.
[3]: There was a lodge formed significantly earlier, in 1858, but it was disbanded.

## B.2 Membership Fees

| | *Minimum* Fee for Degrees | Average Fee for Degrees | Subordinate Lodge Annual Dues | Grand Lodge Annual Dues | Grand Lodge Fee for New Initiates | Year | Reference |
|---|---|---|---|---|---|---|---|
| Arkansas | $25 | | $5 | $0.25 | $3 | 1890 | [1], p. 350 |
| California | $50 | ~$75 | $6-12 | | | 1880 | [2], p. 780 |
| Colorado | $30 | | | $1.50 | $5 | 1890 | [1], p. 424-5 |
| Idaho | $50 | $75-90 | $6-12 | $2 | $3 | 1890 | [1], p. 402 |
| Iowa | $20 | | $2 | | | 1890 | [1], p. 369 |
| Kansas | $30 | | $3-5 | | | 1890 | [1], p. 382 |
| Massachusetts | $25 | | $2-15 | | | 1890 | [1], p. 250 |
| Minnesota | $15 | $30-50 | $2-4 | $1 | $3 | 1890 | [1], p. 354 |
| Montana | $60 | | | | | 1890 | [1], p. 404 |
| New Mexico | $50 | | | | | 1890 | [1], p. 431 |
| North Dakota | $25 | | | | | 1890 | [1], p. 374 |
| Oklahoma\foot | $30 | | $3 | | | 1890 | [1], p. 384 |
| Rhode Island | $25 | $25-200 | $4-25 | | | 1890 | [1], p. 252 |
| Texas | $30 | | | | | 1890 | [1], p. 346 |
| West Virginia | $20 | | | | $3 | 1882 | [3], p. 54, 172 |
| Wyoming | $50 | | | | | 1890 | [1], p. 419 |

Notes:
- There are additional fees for lodge dispensation and charters that are not listed here. This also doesn't include fees for various funds.

[1]: Stillson and Hughan (1890).
[2]: Proceedings of the M. W. Grand Lodge of Ancient, Free and Accepted Masons of the State of California (1880).
[3]: Proceedings of the M.W.Grand Lodge of Ancient, Free and Accepted Masons of the State of West Virginia (1882).

# Appendix C

# Appendices to Chapter 3

## C.1 Proof of Theorem 3.2.1

The proof consists of two parts. The first involves demonstrating consistency of $\widehat{\gamma}$ where either 3.2.1 or 3.2.2 holds. This result is non-standard and detailed below. The asymptotic normality result follows from standard results on sequential estimators (e.g., Newey and McFadden, 1994). The second part of the theorem follows from using Assumptions 3.2.1 or 3.2.2 to simplify the asymptotic covariance matrix for $\widehat{\gamma}$ and checking that this expression coincides with the bound given by Chen, Hong and Tarozzi (2008). These steps are tedious but straightforward and hence regulated to a Web Appendix.

We first consider the case where the density ratio is correctly specified (i.e., Assumption 3.2.1 holds), but the conditional expectation function $q_a(W;\gamma_0)$ may not be not linear in $h(W)$ (i.e., Assumption 3.2.2 does not). In this case we have, under regularity conditions,

$$-\frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)}{\widehat{Q}}\frac{G(k(\widehat{Q})t_i(\widehat{\zeta}_s)'\widehat{\delta}+G^{-1}(\widehat{Q}))}{1-G(k(\widehat{Q})t_i(\widehat{\zeta}_s)'\widehat{\delta}+G^{-1}(\widehat{Q}))}\psi_a(X_i,W_i,\gamma) \xrightarrow{p} -\mathbb{E}\left[\frac{1-D}{Q_0}\frac{p_0(W)}{1-p_0(W)}\psi_a(X,W,\gamma)\right],$$

and, hence, by iterated expectations $\mathbb{E}[m_3(Z,\rho_0,\delta_0,\gamma)]=\mathbb{E}_s[\psi(Z,\gamma)]$, which is uniquely zero at $\gamma=\gamma_0$ by Assumption 3.1.1 ($m_3(Z,\rho,\delta,\gamma)$ is given in (3.23) of the main text). Consistency of $\widehat{\gamma}$ for $\gamma_0$ then follows under standard regularity conditions.

Now consider the case where the density ratio is misspecified, but Assumption 3.2.2 holds. Denote the probability limit of $\widehat{\delta}$ by $\delta_*$ and define $p_*(W)=G(k(Q_0)\,t_i(\zeta_s)'\delta_*+G^{-1}(Q_0))$ and $\Pi_0^*=(\varsigma_0+\Pi_0\zeta_s,\Pi_0)$. Observe that the second step AST moment restriction (equation (3.22) in the main text) implies the population equality

$$t_0=\mathbb{E}\left[\frac{1-D}{1-Q_0}\frac{1-Q_0}{Q_0}\frac{p_*(W)}{1-p_*(W)}t(W,\zeta_s)\right].$$

Note that $\Pi_0^*t_0=\mathbb{E}[\psi_a(X,W,\gamma_0)|\,D=1]$ and $\Pi_0^*t(W,\zeta_s)=\mathbb{E}[\psi_a(X,W,\gamma_0)|\,W]$. Multiplying through by $\Pi_0^*$ and rearranging gives, using Assumption 3.2.2,

$$\mathbb{E}[\psi_a(X,W,\gamma_0)|\,D=1]=\mathbb{E}\left[\frac{1-p_0(X)}{Q_0}\frac{p_*(X)}{1-p_*(X)}\mathbb{E}[\psi_a(X,W,\gamma_0)|\,W]\right]. \tag{C.1}$$

This equality is a consequence of linearity of $\mathbb{E}[\psi_a(W,W,\gamma_0)|\,W]$ in $h(W)$

and the imposition of the moment balancing constraints.

Using this result, Assumption 3.1.1, and iterated expectations we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)\,\varphi_1^{+}(t_i(\widehat{\zeta}_s)'\widehat{\delta},\widehat{Q})}{1-\widehat{Q}}\psi_a\left(X_i,W_i,\gamma\right)\xrightarrow{p}-\mathbb{E}\left[\frac{1-D}{Q_0}\frac{p_*\left(W\right)}{1-p_*\left(W\right)}\psi_a\left(X,W,\gamma\right)\right],$$

and hence $\frac{1}{N}\sum_{i=1}^{N}m_3(Z,\widehat{\rho},\widehat{\delta},\gamma)$ converging in probability to

$$\mathbb{E}\left[\psi_s\left(Y,X,\gamma\right)\middle|\,D=1\right]-\mathbb{E}\left[\frac{1-p_0\left(W\right)}{Q_0}\frac{p_*\left(W\right)}{1-p_*\left(W\right)}\mathbb{E}\left[\psi_a\left(X,W,\gamma\right)\middle|\,W\right]\right].$$

From (C.1) we have the second term on the right-hand-side of the above expression equal to $\mathbb{E}\left[\psi_a\left(X,W,\gamma_0\right)\middle|\,D=1\right]$ at $\gamma=\gamma_0$. The result then follows from the first part of Assumption 3.1.1.