Taxation and State Building Under Diversity

Yusuf Magiya

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2022

Abstract

Taxation and State Building Under Diversity

Yusuf Magiya

The ethnic and religious diversity of the population is often associated with worse state building outcomes, including lower levels of taxation. In this dissertation I investigate how diversity hinders state building and how it shapes the patterns of taxation. The dissertation is structured around two main questions. The first question is: What are the mechanisms through which diversity constrains state building? Building on the fact that periods of state building include increases in the amount of taxes levied on the populations follows the second question that concerns the distributional consequences of the increases in the amount of taxes: Which groups bear the increasing fiscal burdens of an expanding state during periods of state building?

I argue that diversity impedes state building by increasing the costs of the state's investment in fiscal capacity. This is because in more diverse places the different ethnic and religious identities of the population make them more illegible to the state's agents, making it more difficult for the state to acquire knowledge about the population and its economic activities. This illegibility also increases the bargaining power of local intermediaries vis-à-vis the state, which makes investment in fiscal capacity even costlier as these groups often oppose state building. Because it is cheaper to invest in the fiscal capacity of less diverse places, I also argue that the tax burdens of the core/dominant groups in the society, even though they are in power, increase more than the tax burdens of the minorities during periods of state building.

I test these arguments in the context of late-nineteenth and early-twentieth century Ottoman Empire. The main empirical evidence relies on statistical analyses of an original dataset based on my archival work in the Ottoman archives in Istanbul. In addition to this, I use other original and secondary datasets, as well as a close reading and qualitative analysis of correspondences among Ottoman bureaucrats in the Ottoman archives.

Using the local-level fiscal revenue data, I demonstrate that the increases in fiscal revenues during wartime were lower in more diverse areas in the empire, indicating diversity hinders state

building. Using another dataset on the local-level expenses of the state, I find that the state had to invest more in more diverse provinces to be able to extract a unit revenue. This suggests that the costs of investment in fiscal capacity were higher under diversity. In order to provide evidence for the mechanisms I suggest in the argument, I show that the Ottoman State was less successful in successfully completing censuses in more diverse areas, which is consistent with the argument that diverse populations are more illegible to the state. I also utilize a dataset on governor assignments to provide evidence that diversity constrained possible government assignments, potentially decreasing bureaucratic capacity. I complement these quantitative analyses with qualitative analyses of archival documents and evidence from secondary sources.

With these findings, I make three main contributions to the literatures on state building, the politics of taxation, and identity politics. First, I demonstrate that diversity impedes state building, and it does so by rendering populations illegible and making investment in fiscal capacity more costly. Hence, I propose and test a new theory that explains why diversity constrains state building, by bringing together insights from the state building and identity politics literatures. Second, I show that because the members of the core/dominant groups are more legible to the state and investment in fiscal capacity is cheaper where they live, they undergo higher tax burdens of the state building processes compared to the minorities. This indicates a distributive outcome that goes contrary to conventional wisdom where the ruling identity group taxes itself rather than other groups. Finally, finding that war can result in stronger states only under sufficient homogeneity of the population, I underline ethnic and religious diversity as factors that might condition the relationship where war leads to stronger states. This offers one possible explanation why the argument in the wider literature that warfare leads to stronger states is often challenged outside Early Modern Europe, where the populations were less diverse.

Table of Contents

List of Tables

List of Figures

Pronunciation Guide

c/C: "j", as in jam

ç/Ç: "ch", as in chamber

ğ/Ğ: Not usually pronounced, used for lengthening the preceding vowel

ı/I: Written without the dot when lower case. Similar to the "e" in driver

ö/Ö: Similar to the French "eu", or German ö

ş/Ş: "sh', as in shine

ü/Ü: Similar to the French "u", or German ü

Transliteration and Translation

Transliteration of Modern Turkish and Ottoman Turkish words, person or place names, regardless of their language of origin (e.g., Arabic, Farsi, or Turkish) are according to the Modern Turkish orthography. Therefore, for instance, it is spelled as kaza and not qaḍā.

When I am discussing specific words or group of words used in primary or secondary sources, I provide the English translation first and then add the original in parentheses.

Note on Place Names

I use the English names, or names according to the language of the country to which they belong, for those that are outside of the the borders of Republic of Turkey. I also note the Turkish versions of the names in parentheses, unless it has been repeatedly used in the preceding pages. For those places that are located within Turkey now, I use the modern Turkish names.

Glossary

Bureaucratic Capacity: The ability of a bureaucracy to achieve its intended actions

Core/Dominant Group: The group whose language is spoken in the administration, and who belongs to the religious group in power

Düyun-ı Umumiye: The Ottoman Public Debt Administration. An organization established in 1881, under foreign control that aimed to collect revenues from certain tax sources in order to pay the debt on which the Ottoman Empire had defaulted

Ethnolinguistic Fractionalization (ELF): A measure of heterogeneity a population is in terms of ethnolinguistic groups

Ethnoreligious Fractionalization (ERF): A measure of heterogeneity a population is in terms of ethnoreligious groups, taking into account both ethnic and religious identity

Fiscal Capacity: The ability of the state to extract taxes when it has the intention to extract

Fiscal Year: See Rumi year

Governor: The centrally appointed administrator of a province, the highest-level administrative unit in the Ottoman Empire

Hicri Year: The year according to the Hicri Calendar, the Islamic lunar calendar

Independent Sancak: The sancaks (second-level administrative units) that were not under the jurisdiction of any province (first-level administrative unit).

Indirect Rule: The system of rule where some government functions, such as administrative, legal, or fiscal, are delegated to local actors

Kaimmakam: Spelled as *kaymakam* in modern Turkish. Centrally appointed administrator of kaza, the third higest level administrative unit

Kaza: Third highest level administrative unit

Legibility: The extent and the quality of the state's knowledge about the society

Layiha: Report. Often about suggested reforms in an area, among other issues

Millet: Recognized religious community

Muhtar: Head of village or neighborhood

Mutasarrıf: The centrally appointed administrator of a sancak, the second highest level administrative unit

Paşa: The highest level honorary title granted in civil officialdom and military

Province: The highest level administrative unit after 1864

Religious Fractionalization (RF): A measure of heterogeneity a population is in terms of religious groups

Rumi Year: The year according to the Rumi calendar, the solar calendar used for state matters from 1840 on. Based on the Julian calendar

Salname: See yearbook

Sancak: The second highest level administrative unit

Takvim-i Vekayi: The official gazette of the Ottoman Empire, established in 1831

Tanzimat: Literal meaning is reorganization. Denotes the period of reform from 1839 to 1876

Tax Farming: The system where the right to collect revenues from a given tax is awarded to a third party in exchange for a lump-sum payment

Temettüat Registers: Income and property registers prepared as a result of a census at the beginning of the Tanzimat period, in the 1840s

Vali: See governor

Vilayat-ı Sitte: Literally, the six provinces. Consists of six provinces (and in certain periods two others that were later abolished) that had significant Armenian, but also Kurdish populations

Vilayet: See province

Yearbook: Official yearly annals that were published in the nineteenth century. Includes state yearbooks that were published by the central state, but also includes provincial yearbooks that were published by the provincial administrations

Abbreviations

A.}MKT.: Sadaret Mektubi Kalemi

A.}MKT.MHM.: Sadaret Mektubi Mühimme Kalemi Evrakı

A.}MKT.UM.: Sadaret Mektubi Kalemi Umum Vilayat Evrakı

BEO.: Babıali Evrak Odası Evrakı

BOA: Başbakanlık Osmanlı Arşivi

DH.MKT.: Dahiliye Nezareti Mektubi Kalemi

DH.SAİDd.: Dahiliye Nezareti Sicill-i Ahval Defterleri Fihristi

DH.TMIK.: Dahiliye Nezareti Tesri-i Muamelat ve Islahat Komisyonu

ELF: Ethnolinguistic Fractionalization

ERF: Ethnoreligious Fractionalization

ES: Ethnic Similarity

İ.MVL.: İrade Meclis-i Vala

H: Hicri. According to the Islamic lunar calendar

OPDA: Ottoman Public Debt Administration *(Düyun-u Umumiye)*

RF: Religious Fractionalization

ŞD.: Şura-yı Devlet Evrakı

TFR.I.KV: Rumeli Müfettişliği Kosova Evrakı

TFR.I.MN: Rumeli Müfettişliği Manastır Evrakı

TFR.I.SL: Rumeli Müfettişliği Selanik Evrakı

Y.EE.: Yıldız Esas Evrakı

Y.PRK.A.: Yıldız Perakende Evrakı Sadaret Maruzatı

Y.PRK.AZJ.: Yıldız Perakende Evrakı Arzuhal Jurnal

Map of the Ottoman Empire at the End of the Nineteenth Century Showing Administrative Unit Boundaries

Source: Birken (1976)

Chapter 1: Introduction

1.1 The Problem

Fiscal capacity, the ability to raise tax revenues, is a vital component of strong states, and fiscal capacity building is a necessary constituent block of state building. States perform essential functions in providing public goods such as security, the rule of law, education, and healthcare, thereby sustaining economic growth. However, in order to perform these functions effectively, states need funds and therefore to efficiently tax their populations (Schumpeter 1954). States also need tax revenues in order to be successful in warfare, and survive against international and domestic adversaries.

Throughout history, some states have been more successful than others in raising revenues from their populations. The ability to raise revenues has been an important determinant of which states could survive and which vanished (Ertman 1997). It has also been critical for the well-being of their populations as states that could raise more revenues could spend more on public goods. The primary motivation of this project is to explore the sources of the variation in states' ability to collect taxes and succeed in state building.

The literature identifies ethnic and religious diversity as factors that undermine economic and political development (Easterly and Levine 1997; La Porta et al. 1999; Alesina, Devleeschauwer, et al. 2003; Montalvo and Reynal-Querol 2005), including the development of state capabilities (Blaydes 2018), taxation (Alesina, Baqir, and Easterly 1999), and public goods provision (Alesina, Baqir, and Easterly 1999; Habyarimana et al. 2007).

One explanation why states cannot tax efficiently under diversity is decreased demand for redistribution due to differences in preferences. The argument here is that under diversity taxpayers are averse to the benefits of redistribution, such as welfare benefits, made possible by their tax

payments to go to other identity groups, and demand lower taxes (Alesina, Baqir, and Easterly 1999). Even though this explanation is plausible in the modern era where states engage in significant welfare spending which can benefit disadvantaged groups, it cannot explain taxation patterns in contexts where the state spends little on non-security public goods and on welfare, or redistribute resources. Until the emergence of the modern welfare state, most of the expenditures on public goods focused on non-excludable goods such as security. The modern welfare state where the state actively provides large-scale welfare benefits and thus redistributes significant resources from better-off groups to worse-off groups is a relatively recent phenomenon, and was nonexistent before the Bismarckian legislations in Germany starting in the 1880s. It was only adopted by other Western States in the following decades (Fay 1950; Briggs 1961; Kuhnle and Sander 2010). Furthermore, most developing countries started large-scale welfare provision only much later. Similarly, in the historical Middle East welfare and most non-security public goods were often provided by private actors, through the institution of waqf (Kuran 2001; Cansunar 2022).

In spite of the difficulty of extending Alesina et al.'s (1999) explanation regarding taxpayer preferences to other contexts, the relationship between diversity and state capacity in general, and more specifically the relationship between diversity and taxation holds within many different geographical, institutional, and historical contexts with varying welfare state levels, and also in cross-country analyses (Alesina, Baqir, and Easterly 1999; Besley and Persson 2011; Gennaioli and Voth 2015; Blaydes 2018; Charnysh 2022). In this sense, it is worth investigating the mechanisms how diversity can undermine taxation and fiscal capacity. The first and overarching research question that I aim to answer in this study follows these observations: *What are the mechanisms through which diversity constrains state building?*

Periods of state building involve significant increases in the state's fiscal capacity. Policies that can shape the future levels of fiscal capacity, the adoption of new taxes, and increases in the level of taxation often reflect distributive and political conflict across different groups (Mares and Queralt 2015, 2020; Beramendi, Dincecco, and Rogers 2019; Suryanarayan 2021). In this sense, they have consequences regarding which groups pay more and which groups pay less. The state's

attempt to increase revenues by levying higher taxes therefore leads to the political question of how these tax burdens will be distributed across groups. Any group would prefer to pay a lower share of these increasing burdens and make the other groups pay a higher share. This brings me to the second main question in my dissertation about the distributional consequences of state building: *Which groups shoulder the increasing fiscal burdens of an expanding state during periods of state building?*

1.2 Summary of the Argument

My main argument in this study is that the ethnic and religious diversity of the population impedes state building by increasing the costs of the state's investment in fiscal capacity. This is because in more diverse places the different ethnic/religious identities of the population make them more illegible to the state's agents, making it more difficult for the state to acquire knowledge about the population and its economic activities.

I consider two different ways in which a population can be ethnically and religiously diverse. The first is being more *dissimilar* to the *core/dominant* group, meaning speaking a different language and/or belonging to a different religious group than the group in power. The second is being more *heterogeneous*, meaning there are many different groups in the society.[1] Each of these configurations of ethnic/religious diversity makes it more difficult for the state and its agents to obtain information about the population. Hence, each configuration of diversity should, independent of the other configuration of diversity, increase the costs of investment in fiscal capacity, discourage state's investment and constrain state building.

Dissimilarity of the population, where there is a mismatch between the language spoken by the state's agent and the population, makes it more difficult for the agent to obtain necessary information and sustain connections. For agents who are religious outsiders, it can be more difficult to penetrate the networks of the population to acquire information. Without acquiring sufficient information, the state cannot efficiently rule or tax the populations. A heterogeneous population

---

[1] I reintroduce these concepts in Section 2.1.1.

3

presents problems in the same vein because it implies many different dissimilar languages, or religions each of which presents unique problems.

I also argue that merely solving the informational problems are not sufficient to achieve high fiscal capacity. In order to tax the populations efficiently, the state needs to have a high-capacity bureaucracy who can implement policies. I maintain that the ethnic and religious diversity of the population also prevents the state from building stronger bureaucratic capacity because linguistic and religious differences constrain bureaucratic appointment patterns.

Another mechanism in my argument concerns the bargaining power of the intermediaries. I argue that while the diversity and the resulting illegibility of the population makes it more costly for the state's agents to acquire information and rule these populations, it presents an advantage for the local intermediaries because it increases their bargaining power vis-à-vis the state. These intermediaries are often from the local populations, sharing the identities of these populations, and have informational and network advantages compared to state agents. Because these local intermediaries' positions and privileges depended on the state's weakness in their area of influence, they often opposed state building. The higher bargaining power of local intermediaries presents another layer of higher costs of state and fiscal capacity building in diverse places.

Under a more diverse population, the state needs to make higher investments in an area to be able to extract the same amount of tax revenue compared to a less diverse population due to the problem of illegibility and higher investment costs. This implies that any returns to these investments will be higher in less diverse areas. Because of these higher returns, I expect rulers to focus their investments in less diverse areas. As a direct result of these investments, I also expect the increases in fiscal capacity to be higher in less diverse areas compared to more diverse areas. There are a myriad of studies in the literature which maintain that rulers are more likely invest in fiscal capacity building and enhance tax revenues during wartime, under military pressures (e.g., Hintze 1975 [1906]; Brewer 1989; Tilly 1992; Besley and Persson 2010). Theoretically because of this, and empirically to more clearly identify a critical threshold before and after which the ruler's investments should change, I focus on interstate wars. I thus expect to wartime increases in the

state's fiscal capacity to be higher in more similar and more homogeneous places. The patterns where the state invests more in the fiscal capacity of more similar areas and where the tax burdens of these areas increase more than the tax burdens of dissimilar areas imply that the fiscal burdens of state building fall more heavily on the core/dominant groups in the society, who belong to the ruling group. I maintain that this runs contrary to direct and intuitive expectations from a political process, where one might expect the ruling group to tax other groups, and benefit themselves.

1.3 Roadmap

Even though the theory I will outline in this study applies both to ethnolinguistic and religious differences, most of my focus will be on ethnolinguistic diversity for purposes of clarity and brevity. In the empirical sections, I will present quantitative analyses with explanatory variables measuring ethnolinguistic diversity and will evaluate religious diversity only as robustness checks of some key analyses. In the qualitative anaylses sections, most of the focus will, again, be on ethnolinguistic diversity, but I will also discuss some evidence regarding religious diversity.

The remaining part of the dissertation proceeds as follows. In Chapter 2 I introduce a theory of fiscal capacity building under indirect rule and diverse populations. After laying out how diversity and fiscal capacity are conceptualized in this study, I go on to discuss why diversity makes populations illegible and restricts bureaucratic capacity, why it increases the bargaining power of the local intermediaries, and why diversity increases the costs of investment in fiscal capacity through these channels, thereby constraining investment in fiscal capacity. Next, I explain the critical role that wars play, in both the theory and the empirical analyses, as periods when rulers invest in fiscal capacity. Finally, I discuss the Ottoman case and why I expect the theory to be applicable to this case, before introducing the hypotheses and discussing the scope conditions of the theory. Chapter 3 discusses the empirical strategy I adopt in the empirical chapters and introduces the data and the variables I use in each empirical chapter.

Chapters 4 to 7 are empirical chapters. Because I expect fiscal capacity to be dependent on both high legibility and bureaucratic capacity, in two subsequent chapters I test how legibility and

buraucratic capacity are shaped by diversity. In Chapter 4, I test the first part of the argument I propose in Chapter 2 concerning the relationship between diversity and legibility, that expects diversity to increase the illegibility of the population. I demonstrate that the Ottoman State was less likely to be successful in conducting censuses in more diverse areas, indicating that the populations were more illegible in more diverse places. I complement these analyses with a closer examination of correspondences from the Ottoman archives, and evidence from secondary sources which corroborate the finding that it was more difficult to obtain information from the populations that were more diverse.

In Chapter 5 I test how diversity shapes the second critical aspect of fiscal capacity, bureaucratic capacity. Using a dataset of governor assignments in late Ottoman Empire, I find that governors were more likely to be assigned to a province in the region where they are from if their group were the majority group in an ethnolinguistically homogeneous province. I interpret this as evidence that the state needs to assign bureaucrats who possess the necessary linguistic toolkits, knowledge of the culture, and networks to be able to efficiently perform their duties.

Using the same dataset, I also find that while it was easier for the state to assign governors between provinces that had populations with identities that were similar to each other, linguistic differences prevented governor assignment between provinces that had populations with different characteristics. More specifically, the government was more likely to shuffle its bureaucrats between ethnolinguistically homogeneous provinces that was populated by the same ethnolinguistic group compared to shuffling them between ethnolinguistically heterogeneous provinces or shuffling them between homogeneous and heterogeneous provinces. My interpretation of these patterns again relate to the skills of the bureaucrats. This reflects the fact that knowledge and expertise a bureaucrat can acquire in a homogeneous area should be applicable in another homogeneous area where the characteristics of the population are not different. However, such knowledge and expertise cannot be transferable to areas where populations with different characteristics live.

With these findings, I argue I provide evidence that potential governor assignments, where the governors would serve the state more efficiently, are rendered unfeasible due to diversity. This

indicates that diversity can undermine the state's bureaucratic capacity by preventing governor assignments that would otherwise be optimal. I rely on additional evidence from archival and secondary sources which reveal how the diversity of the population undermined building a more skilled and able bureaucracy.

Chapter 6 is the main empirical chapter of this study, where I present the results of analyses with an original local-level fiscal revenue dataset. In the empirical analyses, I focus on changes in fiscal revenues during wartime because it is during wars that rulers are more likely to invest in the state's fiscal capacity. I find that wartime increases in the Ottoman state's fiscal revenues were more limited in more diverse provinces, thereby providing evidence that diversity hinders state building. Furthermore, finding that the wartime increases in fiscal revenues in provinces with a higher proportion of Muslim Turkish speakers were higher, I show that the core/dominant group in the Ottoman Empire shouldered heavier tax burdens of state building.

In the remaining sections of this chapter, I conduct a multitude of robustness checks, evaluate alternative mechanisms, and present evidence from individual-level tax assessment data. In the final section, I discuss qualitative evidence from archival and secondary sources to demonstrate how diversity undermined tax administration and collection, and also undermined military conscription, another aspect of strong state capacity. I also discuss evidence in favor of one mechanism I suggest in the theory, that diversity increases the bargaining power of the intermediaries vis-à-vis the state.

Chapter 7 is the final empirical chapter. It evaluates the second set of hypotheses that expect diversity to increase the costs of investment in fiscal capacity building. I use a dataset on local-level expenditures of the Ottoman State from the fiscal year 1909-10 to calculate an expense-to-revenue ratio, which aims to measure how much the state has to invest in a province to be able to extract a unit revenue. I provide additional evidence from the same dataset where I show that in more diverse provinces the expenditures more heavily relied on items that could sustain state control only in the short run, such as security expenditures, while in less diverse provinces the expenditures were more likely to rely on items that could enhance the state's capacity in the longer run, such as education and infrastructure. I also evaluate some data of bureaucrat salaries, where I provide some evidence

that the Ottoman State needed to pay higher salaries for the administrators who served in more diverse provinces, implying diversity made hiring of bureaucrats more costly.

Chapter 8 concludes the study. In this chapter I first briefly summarize the theoretical argument and main empirical findings. Next, I discuss how the theoretical argument and the findings relate to the literature and contribute to our understanding of the politics of taxation, state building, and identity politics. I also interpret how these findings fit into the general pattern of the decline of empires and the emergence of the more homogeneous national states, before I briefly consider potential policy implications that can be derived from this study.

Chapter 2: A Theory of Fiscal Capacity Building Under Indirect Rule and Diversity

2.1 Defining Concepts

2.1.1 Two Configurations of Diversity

In this study I consider two different configurations of the population's ethnic and religious diversity. The first is what I call *dissimilarity*. A *dissimilar* population has a different ethnic or religious identity compared to the core/dominant group of the society. The core/dominant group speaks the language of the administration and belongs to the religious group in power. I call the opposite of dissimilarity as *similarity*.

The second configuration I consider in this study of diversity is *heterogeneity*. The opposite of heterogeneity is *homogeneity*. Higher heterogeneity implies that there is higher numbers of distinct dissimilar groups in a given population.

These two different measures of diversity aim to capture two different ways in which a population can be diverse in a way that is challenging for the states in terms of information acquisition, for ruling them, and for extracting taxes.

2.1.2 A Conceptualization of Fiscal Capacity

I define *fiscal capacity* of the state as *the ability of the state to extract taxes when it has the intention to extract*. This definition is analogous to what Michael Mann calls the state's infrastructural power, "the capacity of the state to state actually to penetrate civil society, and to implement logistically political decisions throughout the realm" (1984, p.189).

In this study, I conceptualize fiscal capacity as a combination of the state's informational capacity (or using the term that I will prefer during the rest of this study, the *legibility* of the society to the state). In order to have high fiscal capacity and to be able to tax efficiently, a state needs to have sufficient informational capacity as well as sufficient administrative capacity.

Legibility can be briefly defined as the extent and the quality of the state's knowledge about the society. A state that has rendered a society more legible possesses more and better standardized information about the society, which is more easily understood by any agent of the state (Scott 1998; Zhang and Lee 2020). Following Huber and McCarty, I define bureaucratic capacity as the bureaucrats' *"ability to accomplish intended actions"* (2004, p.481).

I argue that it is not possible for a state to accomplish high fiscal capacity without achieving both high levels of legibility (Christensen and Garfias 2021; Hau, Peres-Cajias, and Soifer 2021), and bureaucratic capacity (Kasara and Suryanarayan 2015). In Levi's words, "Bureaucracy requires considerable personnel and efficient dissemination of quality information" (1988, p.178). Without legibility, lacking sufficient high-quality standardized information about the population and its economic activities, the state would face challenges extracting taxes from regardless of how able its bureaucrats are.

The process of state building includes increases in the size or the rationalization of the bureaucracy (Fukuyama 2011, p.332), or a more efficient one (Brewer 1989). Without an able bureaucracy that can implement what tax policy the state aims, and without able agents who can go out to the field and extract the taxes, it should not matter how much the state knows about the society.

2.2 The Link Between Diversity and Fiscal Capacity Building

2.2.1 Diversity and Legibility

The state can be perceived as seeking to maximize its revenues given certain constraints (Brennan and Buchanan 1980; Levi 1988). One essential factor that can limit the state's ability to tax and to rule is the knowledge that the state possesses about the society and its economic activities (Brewer 1989, p.221; Scott 1998; White 2004; Mayshar, Moav, and Neeman 2017). The state can effec-

tively tax as long as it knows whom and what there is to be taxed in the society and the economy. It can effectively rule as long as it has sufficient funds and has enough knowledge about its subjects. However, it is seldom straightforward for the state to acquire information from a population, as people are rarely willing to share information in an easy manner with the state agents (Weber 1976, p.44). If the people share information about themselves, it can help the state levy heavier taxes on the them and they can be conscripted into national armies.

State building is often regarded as a struggle between the state and the society, (Scott 1998, 2009), where the structure of the society affects the state's capabilities (Migdal 1988). Ethnic and religious diversity are two factors regarding the society's structure that can affect how much the state knows about the society and also its capabilities (Scott 1998; Blaydes 2018; Charnysh 2022).

James C. Scott writes that "a unique language represents a formidable obstacle to state knowledge" (1998, p.72). According to him, "the great cultural barrier imposed by a separate language is perhaps the most effective guarantee that a social world, easily accessible to insiders, will remain opaque to outsiders" (p.72) and hence will be *illegible*.

The different languages different ethnic groups speak, their different customs, different economic activities different groups specialize in make each one a separate group for which there is unique information not readily available to any outsider.[1] Each group that has a different ethnic identity than the state's dominant ethnic identity forms one "unique language" in Scott's (1998, p.72) terms and higher ethnic heterogeneity increases the number of "unique languages", rendering the populations more illegible to outsiders, who do not speak that language and are unfamiliar with the culture. Therefore, ethnic dissimilarity and ethnic heterogeneity make the population in a given area and their economic activity less legible to the state, which makes it more difficult for the state to learn about, monitor and project its power on such populations (Blaydes 2018).

We can expect similar patterns regarding religious diversity. If the bureaucrats are assigned to places where their co-religionists live, they can more easily penetrate their networks since it is easier for them to build trust (Livny 2020). In addition to this, the state can acquire more

---

[1]For example, a group can specialize in the production of a specific product, or focus on an economic activity which the outsiders know little about.

11

information about the groups who share the state's dominant religion since the state likely has denser connections with the clerics and leaders of such communities.

That heterogeneity and mismatch between a population's and the center's ethnic or religious identity *(dissimilarity)* renders a society more illegible to the state and that these make it more costly to obtain information about these populations, monitor and control them should also prevent the state from achieving fiscal centralization and increase tax revenues. This is likely to hold especially under regimes of indirect rule—such as empires—where the central authority relies on intermediaries for administration, especially tax collection.

States aim to make local practices more legible to its officials (Scott 1998). This is a key aspect of state's control of the periphery and also shapes how much revenue it can extract. In order to think about the mechanism through which the illegibility of a population undermines fiscal capacity building, we can rely on Levi's (1988) framework, where relative bargaining power of the rulers and transaction costs are two constraints on the state's fiscal capacity.[2] These transaction costs can include the costs of monitoring the population's activities (Blaydes 2018), obtaining information about revenue sources and constituent behavior, enforcing compliance (Levi 1988), obtaining information about the type of product to be taxed (Coşgel 2005), or merely the costs of communication and interactions between the taxpayer and the state agent (Zhang and Lee 2020).

In diverse contexts the transaction costs of communicating with the taxpayer, counting, monitoring and controlling a population, and obtaining other necessary information for taxation are higher. This makes the investments to increase fiscal capacity more costly. Moreover, the high transaction costs increase the premium on local intermediaries, who have better information about the local context and have better connections. This decreases the ruler's bargaining power vis-à-vis local intermediaries. The intermediaries possess critical information and connections that the state agents lack. Consequently, its lower bargaining power makes it more difficult for the state to eliminate these intermediaries, centralize tax collection and hence increase fiscal capacity. This especially is a problem for contexts with indirect rule, such as empires where states initially have

---

[2]She also discusses discount rates as the third constraint, which I do not consider in my theoretical framework.

little knowledge about its populations, and for rural contexts with agricultural economies, which limit the state's knowledge, reach and control.

In order to obtain information about the population and render it legible, states can rely on censuses (Scott 1998; Lee and Zhang 2017; Brambor et al. 2020). With reliable censuses, any agent of the state, whether looking at the census while sitting at an office desk at the capital or while in the field at the periphery, can learn about the population. This information can be critical for tax collection purposes. They can communicate to the reader certain factors such as how many people live in certain localities and their economic activities. What is problematic regarding diversity for the administration of censuses is that dissimilar ethnic and religious identities and heterogeneity increase the costs of administering censuses and therefore the costs of obtaining information. For an outsider who does not speak the language, unfamiliar with local customs and lacking the necessary connections, locating and accessing the settlements, reaching the populations there, counting these populations and obtaining the relevant information about them is difficult.

Similar mechanisms are at play regarding tax assessment and collection. Agents who are not familiar with the region do not possess the necessary knowledge of what product is grown where and when it is harvested, which are critical for the collection of agricultural taxes. They also lack adequate knowledge of the geography, which can make it difficult for them to navigate. They neither have the proper means of communication with local figures, nor with the taxpayers and it is more difficult for them to control lower-level local (non-state) actors who can speak the language and know the area, because they have no proper connections with such actors. Establishing connections with lower-level local actors are often key to controlling an area also because they are likely to possess critical knowledge and connections with the local population and have earned some combination of trust and reputation (good or bad) which can make it more likely for them to extract taxes compared to an outsider. Furthermore, these local actors should be likely to stay loyal to already or previously dominant intermediaries because of their already well-established connections.

2.2.2 Diversity and Bureaucratic Capacity

The diversity of the population can also affect the state's bureaucratic capacity through two inter-related channels. First, diversity can restrict bureaucratic appointment patterns. This is especially problematic for linguistic differences. A mismatch between a state official's and the population's language is likely to undermine a state agent's ability to communicate with the locals, acquire information about them, and perform their duties (Weber 1976; Blaydes 2018). In such a case, not every agent can be assigned to the position where they can serve the state best given their skills, but the state needs to appoint agents to certain positions given their linguistic skillset.

The second channel through which diversity can affect bureaucratic capacity is having to appoint local agents with roots and strong networks in the area, which invites possible patronage relationships. The central state can rely on such agents as a solution to the problems that any centrally-appointed agent can face. This can be advantageous in the sense that in the short run they can rule better compared to a centrally-appointed agent who may not be speaking the language and familiar with the customs in the area. However, because diversity also restricts bureaucratic appointment patters, as I discussed above, agents with local roots can undermine the state's bureaucratic capacity in the longer run as they can build patronage relationships with the local population to the detriment of the central state. If the state needs to assign these bureaucrats elsewhere in order to prevent these patronage relationships, with the lack of alternatives to fill these positions, this will also undermine the state's capacity. Local elites when appointed as bureaucrats are less accountable to their superiors in the national bureaucracy (Soifer 2015). Indeed, to what extent local agents take up bureaucratic positions can be used as an indication of state capacity (Köksal 2002).

2.2.3 Local Intermediaries and Bargaining Power

To better explain why diversity and the illegibility it causes increase the premium on local intermediaries in rural and agricultural contexts with indirect rule, it is a good idea to consider how governance and taxation in such contexts usually worked. Under indirect rule the central state

relied on local intermediaries to rule the populations. This was a very common form of rule in the empires in the era before what Tilly (1992) calls nationalization, before empires and city-states converged towards 'national states'. Empires contained diverse populations. They usually were multi-ethnic and multi-confessional. However, "Almost universally [empires] were overtly or tacitly the empires *of* a particular people or ethnos. The military might and political dominance of that people/ethnos was the cornerstone of the empire. The empire's rulers and most of the elite came from it." (Gat and Yakobson 2013, p.111), indicating that in spite of the diverse populations in empires, the core/dominant group usually was of a certain identity group.

With a diverse population, and the ruling group belonging to a specific identity, populations in empires were naturally illegible to the state's agents, who often came from the core/dominant group. Such illegibility made it more difficult to rule directly from the center, and made indirect rule a better strategy for the state. They needed the cooperation of the local forces in order to project their rule. The illegibility which made it more difficult for the state to obtain information and rule over subjects compelled states to delegate administration and tax collection to intermediaries who possessed local knowledge and networks. Yet, when the intermediaries were involved in the activities in ruling and taxing, they used their position to benefit themselves to the detriment of the central state (Doumani 1994).

Tax farming was a common way for the rulers to avoid risk and ensure a steady flow of income in indirect rule (Levi 1988; Johnson and Koyama 2014). By delegating tax collection to tax farmers, rulers could ensure some income from the initial payment from the tax farmers regardless of the result of the tax collection attempts and did not have to undertake the costs of tax collection. Especially farming the administratively complex taxes or those that were more difficult to collect was preferable for the rulers (Brewer 1989). The downside of tax farming was that a substantial portion of potential revenues ended up in the pockets of these tax farmers (Özbek 2015). In the era of nationalization that Tilly (1992) describes, those states which could manage to converge towards national states were particularly advantaged in warfare. During the transformation from empires to national states, and from indirect to direct rule, states "took over the direct operation of

the fiscal apparatus, drastically curtailing the involvement of independent contractors" (Tilly 1992, p.29). Nonetheless, the transformation from indirect rule to direct rule was costly for the state and this cost depended on the bargaining power of local intermediaries (Garfias and Sellars 2021).

State building projects are argued to be more likely to be successful with outside-appointed agents compared to the appointment of local elites (Soifer 2015). This is because such elites, including local intermediaries are likely to be losers of state building (Soifer 2015; Yaycıoğlu 2016; Sezer Feyzioğlu 2017). They would in a way lose their jobs if the state became strong enough in their area of influence, being able to assign agents from the center to perform the functions the local intermediaries were needed for.

2.2.4   Higher Costs of Acquiring Information and Capacity Building Under Diversity

The state can be analyzed as an actor that engages in a cost-benefit analysis and decides on the optimal level of investment in fiscal capacity given its costs and benefits (Herbst 2000; Besley and Persson 2010; Gennaioli and Voth 2015).

In light of the discussion in the previous sections, I expect it to be more expensive to count the populations and also to centralize tax collection in dissimilar and heterogeneous areas. In order to centralize the collection of a given amount of tax revenue, the required amount of investment (both in the short- and long-term) should be higher in more dissimilar and heterogeneous areas compared to similar and homogeneous areas. In dissimilar areas it should be more expensive because the state will need to send bureaucrats who speak the language of and share a similar identity with the local population where dissimilar people live. State bureaucracies often use the language of the dominant ethnic group and most of the state officials speak this language (Blaydes 2018; Zhang and Lee 2020). Finding bureaucrats who can speak the language of the local populations and establish connections in such places is often more difficult (BOA.BEO. 1968/147542). In heterogeneous areas it should be more expensive because heterogeneity implies multiple dissimilar ethnic and religious groups and the state will need to hire more bureaucrats and therefore has to spend more money to extract a given amount of revenue from a given area (this is a strategy that can be adopted

16

in the shorter term) and invest in education that will contribute to the bureaucrats' skills, including language skills that will help them communicate and extract resources from different ethnic groups (this is a strategy that can work only in the longer term). Another long-term investment for the state to pursue can be to invest in communication or transportation infrastructure, which can make information acquisition cheaper.

A simple hypothetical example in Figure 2.1 below may help visually illustrate how in more diverse contexts it can be more costly for the state to tax the populations due to legibility problems, or lack of bureaucratic capacity. Let us assume three hypothetical contexts, with four districts (consisting of four cells that are demarcated with bolded lines) and four subdistricts (each cell), that are similar in all aspects except that in the first example the population exclusively consists of ethnic group A (example I), in the second all four districts are homogeneous within each other but the population in each district are from a different ethnic group, which I call groups A, B, C and D (example II) and in the final example populations in each subdistrict within a given district are from a different ethnic group (example III). Let us further assume that a given tax collector's performance (i.e. ability to extract taxes) becomes worse if assigned to cover a larger area (which should be obvious since the time and resources this collector can spend on a given area declines if assigned to cover a larger area).

Figure 2.1: Demonstration of Three Hypothetical Cases of Ethnic Composition

I- Homogeneous and Similar   II- Homogeneous within Subdistricts but Mostly Dissimilar within the District   III- Heterogeneous and Mostly Dissimilar

Let us say there are four agents at the state's disposal and each of them will be assigned to collecting the taxes of four subdistricts. Let us also assume that this is a state where the dominant ethnic group is A, and an average agent is from ethnic group A, speaking this group's language.

17

Finally, let us assume that this language is the 'official language' used in public matters. In example I, centralization of tax collection should be the easiest. Any four regular agents can be sent to any of these districts and can be responsible for tax collection in all four subdistricts within this district. Example II is a more difficult hypothetical case from the state's perspective. Here, one agent can be sent to each district and will be responsible for tax collection in all four subdistricts in a district. However, the problem here is that each district has populations of a different ethnic group. This is why ideally the state should send agents that match each district's identity. Unless the state has already invested in the education and hiring of agents who speak languages B, C, and D; effective tax collection in these districts will be more difficult, with most agents speaking language A.

Example III presents the most challenging case. Here, each district has populations from each ethnic group. Not only the mismatch between the state's ethnic identity and some of the population's ethnic identity, but also the ethnic heterogeneity within each district is a problem. Even if the state has already invested in recruiting agents from groups A, B, C and D, tax collection in this case is very difficult unless all of these agents speak all four languages. Comparing the amount of travel an agent who speaks a single language, and is assigned to collect taxes from subdistricts where his language is spoken, needs to do in this case to the travel the same agent in cases I and II needs to do should be helpful in understanding this point. While in cases I and II this agent needs to travel through a much smaller area to cover four subdistricts (the agent only needs to travel within a given district), in case iii the same agent will need to travel through a much larger area to cover the same 4 subdistricts. This can undermine the efficiency of the agent's collection, or the state will have to assign more agents. Each of these options make this case costlier than the other cases. The within-district heterogeneity is a problem for all districts in case III.

For these reasons, I expect tax administration and tax collection to be more costly where the populations are more is both dissimilar and heterogeneous, and I expect them to increase the costs of these independently of each other. These costs should determine where the ruler will invest when there is a need for more revenues. Aiming to achieve maximum returns from the investment in fiscal capacity, the rulers should be more likely to invest in the fiscal capacity of more homogeneous

and similar areas because the returns to investment in such areas will be higher compared to more heterogeneous and dissimilar areas, where the same amount of investment will have lower returns.

2.2.5   How War Changes the Ruler's Calculus

During peacetime there is already an established tax collection system which is optimal given the peacetime revenue needs of the ruler. With the war acting as a shock, there is an equilibrium shift where the ruler needs more revenues. For this reason, wars usually compel rulers to invest in the state's fiscal capacity (Tilly 1992; Besley and Persson 2010). Many studies find that wars, directly or under some conditions, enhance the state's fiscal capacity in the short (Karaman and Pamuk 2013) and in the longer term (Queralt 2019). Although the advent of war does not necessarily affect the costs of investment in fiscal capacity, or cheapness of tax collection, I argue that war is critical because winning or losing a war determines the ruler's future benefits, and the probability of winning or losing a war often depends on how much a ruler invests in warfare. The more resources the ruler can extract from the population, the more they can invest in the war and increase their probability of winning the war.

The ruler's future benefits can be affected by war's result via win or loss of territory and hence win or loss of future revenue from this territory, or via affecting the ruler's survival.[3] Therefore, the direct returns from an investment in fiscal capacity may be equal during wartime and peacetime, but the indirect benefits through increasing the probability of winning the war make the overall returns from an investment higher during wartime. This is why we may see a different investment pattern in fiscal capacity during wartime, compared to the ongoing investment patterns during peacetime. However, the relationship between expected benefits from winning the war and spending in war are probably not linearly related. Spending very little against a very strong opponent in war should not be any different than not spending at all in terms of the probability of winning the war.[4]

---

[3]A ruler's survival in office and his literal survival can be affected by war's outcomes due to occupation by the foreign power, but also happen by the economic and political turbulence losing a war creates and the ruler being blamed for this (Croco and Weeks 2016).

[4]For example, spending resources on an army of 1,000 soldiers or spending resources on an army of 5,000 soldiers against a rival army of 100,000 soldiers should be equally futile.

Figure 2.2: Contest Success Function According to Levels of Spending on War
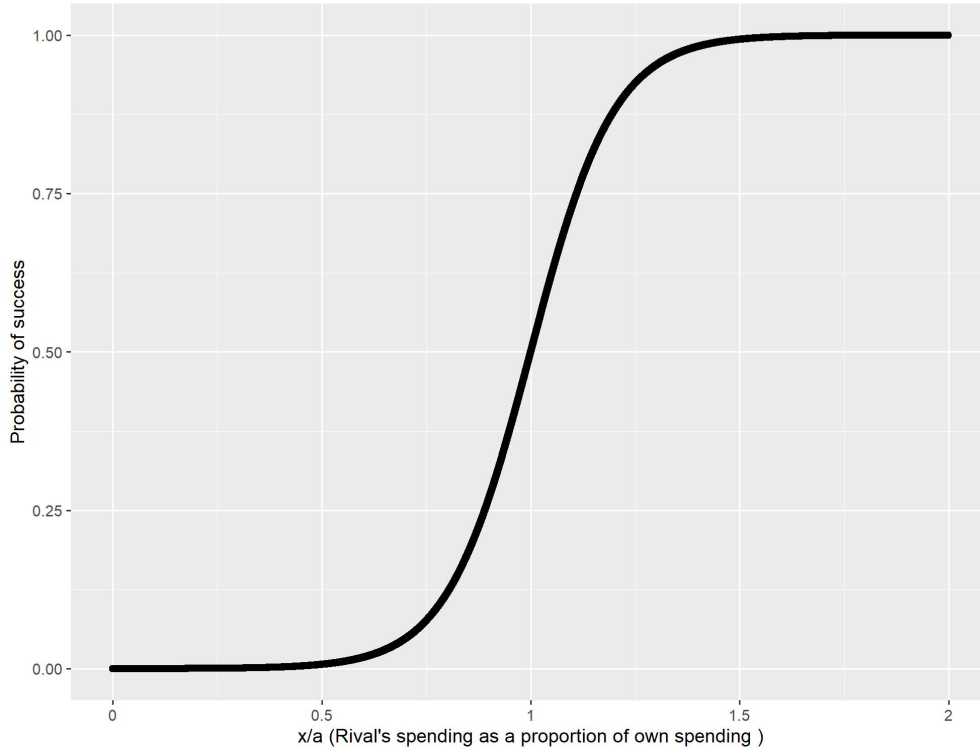
Note: Plotted for $\alpha=10$

We can see this more clearly if we consider the probability that the ruler wins the war as a result of a contest success function, following a strand of literature that analyzes the outcome of violent conflict (Skaperdas 1992; Hirshleifer 1995). Based on Margaret Levi's expectations, who writes that "war making sometimes took on an S-shaped function" (1988, p.106, fn. 13) of the spending in war, I will assume a contest success function of the form:

$$p = \frac{e^x}{e^x + e^\alpha}$$

where p is the probability that the army wins the war, x is the amount of expenditures the ruler makes in the war and α is the amount of expenditures the rival country makes in the war. Figure 2.2 plots how the probability to win the war, p changes with $\frac{x}{\alpha}$, the proportion of its rival's war expenditures to the state's war expenditures.

The probability of success does not change much with higher spending at low levels of $\frac{x}{\alpha}$. However, as the spending approaches the rival's spending from the left hand side, even a small amount of extra spending in war can result in drastic increases in the probability of success. I argue that instead of investing its resources in war at low levels of $\frac{x}{\alpha}$, the state can invest these resources in extracting higher revenues, and with the larger resources they can obtain, their spending in war can make a much larger impact in its probability of winning the war and in its future benefits.

One important aspect of war increasing fiscal capacity, if the legibility problem has been solved, is that it not only causes short-term and temporary increases in fiscal revenues, but can also help keep fiscal revenues higher even after war is over, constituting a 'ratchet effect'. Once the ruler has invested in a given area, solved the legibility issues here and has centralized tax collection, the costs of continuing with this arrangement are lower. In sum, this ratchet effect occurs because, in Tilly's words, the "wartime increase in state power gives officials new capacity to extract resources" (1992, p.89).

2.3 The Case Context and the Hypotheses

2.3.1 How Diversity Hindered Revenue Extraction in the Ottoman Empire

The Ottoman Empire originated in what is today North-Western Turkey, and later expanded into the Balkans and Central Anatolia, before further expanding towards territories in Central/Eastern Europe, Middle East, and North Africa. In the Balkan provinces, the Empire had inherited a difficult-to-rule heterogeneous population from the Byzantine Empire, who also had had troubles ruling these heterogeneous populations (Karpat 1982, p.148). In response to the challenges of ruling diverse populations speaking different languages and of different faiths, the Ottoman State designed the Millet system, which according to Karpat (1985b) aimed for the bureaucratization of the confessional religious structures. The subgroups of Millets were headed by lower clergy. These clergy were from the ethnic group and spoke the language of the community between which and the state it was an intermediary (Karpat 1985b).

After the "Military Revolution" that took place roughly between mid-sixteenth and seventeenth centuries (Roberts 1956; Parker 1988), wars became longer and more costly, and required states to rely on standing armies. The Ottoman Empire had to increase its revenues, or face the risk of perishing in war like many other examples in the Europan State System (Downing 1992; Tilly 1992; Ertman 1997). Mehmet Genç argued that the principle of fiscalism, was one of the principles that shaped Ottoman economic policy making. This principle aimed to maximize the income of the Ottoman State treasury. In spite of this, the Ottoman Empire ran into challenges in trying to increase its revenues until the end of its existence. Much of the reforms in the Empire during the nineteenth century aimed at reforming the bureaucracy and tax collection system in order to be able to collect sufficient revenues to counter the military strength of its rivals.

With the Imperial edict of 1839, Tanzimat Fermanı, the Ottoman Sultan declared that tax farming would be abolished. By eliminating local intermediaries who often undertook the duty of tax farming and centralizing tax collection, the state aimed to increase the revenues of the central treasury. Tax farmers and their financiers pocketed a substantial portion of the potential revenue that could have ended up at the central treasury.[5]

The state initially abolished tax farming in some regions of the empire and appointed certain bureaucrats named muhassıl-ı emval with the task of allocation, assessment and collection of taxes (Pamuk 2016; Özbek 2015, p.48). Yet, the tax revenues of the central state fell drastically after these officials undertook these tasks, due to the problems they encountered in tax collection (Pamuk 2016, pp.200-1). At the basis of the problems in tax collection lay the weakness of the state in the provinces of the empire (Pamuk 2005, p.201). These weaknesses included the inadequate local knowledge of the centrally appointed agents, and lack of cooperation from previous tax collectors and previous provincial administrators (Köksal 2002). In other words, at least part of the reason was the illegibility of the society and its economic activities to the state, which increased

---

[5]Late-19th century tax administrators in the Ottoman Empire estimated that around 15-20% of the agricultural tithe revenue was lost to tax farmers (Özbek 2015, p.29). This was not the only loss of revenue due to tax farming. Short-term tax farmers often overtax the populations due to their short time horizons and harm agricultural production. The amount of revenue lost to intermediaries could be higher before the early nineteenth century centralizing reforms. Çizakça (2013, p.255) estimates that after the introduction of lifetime tax farming *(malikane)* in 1695, only 24 guruşes reached the Ottoman treasury per each 100 guruşes collected from the public.

the transaction costs of tax collection (Özbek 2015) and the indispensable nature of the local intermediaries, which increased their bargaining power. The costs of direct administration of the taxes were indeed staggeringly high. Çadırcı (1989, pp.237-8) cites the 238. issue of the Official Gazette of the Ottoman Empire in this era, *Takvim-i Vekayi*, published on February 14 1842, where it was reported that the muhassıls in charge of the administration of taxes had to hire many agents to collect agricultural tithe during the harvest season and the wages of these agents amounted to half of the total revenue from this tax. The institutional innovation aiming to establish direct collection of taxes having failed, the central government reinstituted tax farming in 1842 and continued relying heavily on the local intermediaries for the collection of agricultural tithe.[6]

Replacing tax farmers was difficult also because the empire suffered from a chronic lack of a well-educated workforce (Akyıldız 2004; Barkey 2008). It was costly for the Ottoman state to invest in a workforce that could be appointed from the center to any given region in the Empire with the the different languages, religions, customs, networks and economic activities of different groups in different parts of the country. Comparing the Ottoman case with 17th century England can help illuminate how diversity can undermine building a centralized bureaucratic apparatus. Brewer examines how during the reign of James II of England in the 1680s the foundations of a strong tax administration were laid:

> "The link between periphery and centre was tightened. Customs commissioners began regular circuits of inspection, checking officers' performance in the field. Similarly the excise commissioners and subordinate officers with the titles of general riders and general supervisors toured the countryside inspecting the work of their gaugers. The Excise introduced standardized instruction for their employees according to 'the Method'. Officers were brought to London for instruction and, when properly 'Methodized', were sent back to the provinces to teach their fellow-officers the most effective means to gauge and measure taxable goods." (1989, p.94).

---

[6]Despite other attempts of Ottoman administrators to abolish tax farming, and centralizing tithe collection in some areas and in some periods, these attempts were never permanently successful during the lifetime of the empire. It was only in the much more homogeneous Turkish Republic in 1925 that it was abolished (Özbek 2010). In 1943, during World War II, tithe was reintroduced under a new name, *Toprak Mahsulleri Vergisi*, which was abolished in 1946.

While this option worked for the government of James II in the 1680s, it was less likely to be viable for the Ottoman rulers. England at the time was probably ethnolinguistically more homogeneous than the Ottoman Empire. It was possible to send the customs commissioners to check the officers' performance, 'the excise commissioners and subordinate officers' could tour the countryside and inspect the gaugers, and introducing a standardized instruction and bringing officers to the capital to instruct them before sending them to the provinces, who would then teach other officers was not that difficult. These were possible because in England most—if not all—officers probably spoke English, which—probably for most—was the language they needed to administer taxation in the field. The diversity of the languages in the Ottoman Empire should be less likely to allow such an increase in administrative capacity and an ensuing increase in fiscal capacity. Indeed, the Ottoman state was often compelled to find officials who could speak the local languages when sending administrators to the provinces (BOA.TFR.I.A. 36/35; BOA.TFR.I.MN. 138/1370). Before the census of 1881-93 in the empire, a debate among high ranking officials regarding the administration of the census can provide some insight on how different languages troubled the state.

"During the debates in the Şura-yı Devlet (council of state), the General Committee suggested that since some non-Muslim community leaders could not speak Turkish and consequently faced difficulty in filing the birth certificates, the population officers sent to these areas should be selected from those who spoke the native languages. In other words, instead of compelling the citizens to learn Turkish, the language of the administration, the Ottoman government sought at this date to teach its own officials the regional languages, a custom long in practice." (Karpat 1985a, p.32, fn. 38).

Such language barriers should pose a disadvantage for the state for several reasons, as I discussed in the preceding sections. They narrow the pool of officials the state can send to the provinces, make it more difficult for the state to freely assign a given official to any vacant position, or to could promote a well-performing official to any position where he could serve the state better.

Linguistic barriers undermine such necessities of better bureaucratic capacity. Consequently, they can also undermine the fiscal capacity of the state.

The Ottoman Economy heavily relied on agriculture in the nineteenth century and a vast portion of the state's revenue came from agricultural tithe *(aşar)*, for which legibility was a critical phenomenon.[7] Tithe was a difficult tax to administer and collect because agriculture, by its nature, is mostly a rural economic activity. What is worse, few Ottoman agricultural producers produced for the market and engaged in subsistence agriculture, which made their economic activity less legible to the state.[8] It was more difficult for an outsider who does not speak the language, unfamiliar with the geography and the type of the products grown in certain localities to administer and collect this tax. Tithe was often taken in kind (Özbek 2010) and the assessment of the output had to be done right after the harvest, before the product was moved elsewhere. The fact that different types of products were harvested in different times of the year presented a significant obstacle to outsiders, who were less likely to know what product was grown where and when in this specific climate a given product's harvest time was. It was even more difficult if the same farmer grew different products that were harvested in different times (Özbek 2010). The "narrow timeframe of revenue intake" (Anscombe 1997, p.59) from tithe assessed on agricultural output was a constant challenge to the Ottoman State. Any delay in the assessment and collection of in-kind taxes such as tithe could lead to the peasants selling all their product, leaving nothing to the state treasury (Doumani 1995, pp. 108-9).

A safer alternative to output taxes on agricultural produce could be to apply input taxes (Coşgel 2005), where taxes could be assessed based on the input, such as the land area, or number of trees, rather than assessing the amount of output. However, it is more difficult to increase tax revenues with input taxes; for example, even if output per unit input increases over time, there is no way

---

[7]For instance, tithe income constituted around 31% of the total revenues in the fiscal year of 1860/61 and 44% in 1876 (Özbek 2010, p.55).

[8]I also expect illegibility to decrease the efficiency of other types of tax collection, such as taxes on transactions like trade or taxes that were mostly applied to urban areas. However, among all groups the most illegible group to the state were villagers living in remote areas and the most illegible activity for tax purposes was subsistence agriculture.

to prove this on the part of the tax official and therefore it is less likely for the state to reap the benefits of higher production.

This is why the rulers had to rely on local intermediaries and tax farmers for the taxation of agriculture. Tax farming helped raise revenues from "constituencies that the ruler understood poorly" (Kuran 2011, p.134) through intermediaries who understood them much better. The intermediaries had extensive knowledge about the locality, had established dense networks among the population and therefore did not experience the difficulties that centrally appointed agents faced. Özbek (2015, p.97) notes that those who got the tax farming contracts were the already well-established notables *(eşraf)* in the provinces. They were often indispensable due to the illegibility of the society to outsiders, in this case the centrally appointed agents. Most of the intermediaries were insiders with much better knowledge of the society. They also had much denser connections within these local communities, which enabled them to extract higher resources compared to any outsider appointed by the center. Nevertheless, such a reliance on them increased their bargaining power and they could use such bargaining power to increase the portion of the revenue they could acquire themselves, thereby decreasing the state's revenues.

2.3.2   How Fiscal Revenues Increased in the Ottoman Empire

The Ottoman state's wartime fiscal revenues may have increased in several ways. For the taxes that were directly administered, via the state's own agents, the problem was the transaction costs of tax collection. When it was in need of funds and as long as the costs were not high enough, the state could invest in fiscal capacity by spending to learn more about the population and its economic activities, and hiring more competent agents. In these cases, I expect the costs of necessary investment increase as diversity increases. Hence, the state is more likely to invest in fiscal capacity under more homogeneity and similarity.

For the taxes that were farmed, the problem was both transaction costs of tax collection and the higher bargaining power of the intermediaries due to the inherent advantages they possessed in local knowledge and networks. Where intermediaries were involved in tax collection, the revenues

could have increased via two paths. Firstly, with the urgency of the resource extraction and the higher value of the future flow of taxes, the ruler could eliminate the intermediary from the process and start collecting the taxes directly, a process I call fiscal centralization. Having eliminated the intermediary, the state could lay claim on all the money that could be collected, not having to share a portion of it with the intermediary. As long as the necessary costs of the investment were low enough, the state could choose this path. However, the elimination of the intermediaries was not always necessary to increase the state's fiscal revenues for the taxes that were farmed. I argue that the second path that could increase the state's fiscal revenues was the *possibility* of fiscal centralization by the state; when the state's bargaining power against the tax farmers increased and the state could sell the farming contract for a higher price. In order to better explain this mechanism, it may be a good idea to take a look at how tax farming contracts were allocated. I will focus on the agricultural tithe, the main source of the state's income the during this period.

The system of farming the agricultural tithe was very heterogeneous across time and space in late Ottoman Empire. The allocation methods changed very often and not all changes applied to every region. To give a crude summary, Özbek (2018) identifies two ways of allocating rental tax farming contracts of the agricultural tithe, fixed price *(maktu)* and ascending price auction *(müza-yede)*. In the fixed price system, the government would determine a fixed price for the contract. In the ascending price contract, the bidder with the highest bid would win. Those contracts that no one bought were often administered via wage contracts *(emaneten idare)*, which means that at times they were directly administered via agents assigned from the center and at other times the duty to collect these taxes were given to the village communities *(ahali-i kura)*.[9]

The state's net profit (its tax revenue minus expenditures to collect these taxes in this unit, such as wages paid to the bureaucrats and other costs of assessing the taxes) it could obtain from

---

[9]Although the name suggests that the duty to collect taxes were assigned to the village communities as a whole, in practice it was often the notable members of these communities such as village heads (muhtar) or clerics who undertook this duty (Özbek 2015, p.89). Under low capacity, allocating tax collection duties to members of the communities also created problems. Akyıldız (2004, p.204) notes that many villagers complained to the state that their village heads engaged in corrupt practices during tax collection such as overtaxation, or not giving receipts in return for tax payments. Lacking capacity, the state had no means of learning what was really going on and had to accept the village head's statement regarding the tax revenue.

directly administering the tithe of a given unit should depend on the population's legibility. The more illegible the population in a unit is, less return from a given investment the state should obtain in tax administration. The net profit that the state could obtain from administering the tithe here also forms a baseline revenue for the state. We can assume that the state would farm the tithe in a unit if a tax farmer offered a higher price than its baseline revenue. The expenditure of a tax farmer with a better knowledge of the locality and population should be lower than the expenditure of the state. This is why in many cases the contract should have been sold for a higher price than the state's baseline revenue.

During wartime, when the ruler is in urgent need of increasing revenues, the benefit of a given amount of revenue should be higher for the ruler compared to peacetime (as I have argued in Section 2.2.5 above). This makes it more likely for the state to mobilize resources to centralize tithe administration and therefore the potential baseline revenue increases. This can allow the state to ask for a higher price in the fixed price contract system and can also drive the price up in the ascending price auction system. Thus, whether or not a farmer buys the contract and also whether or not the state centralized tax collection in this unit, the income of the state should increase. This mechanism should be more likely to happen in more homogeneous and ethnically similar places and less likely to happen in more heterogeneous and ethnically dissimilar places. Heterogeneity and dissimilar ethnic groups, as I have argued, increase the amount of investment the state has to make in overcoming the barrier of illegibility in this area and therefore the wartime increases the baseline revenue in these places should be lower, it is even likely that the baseline revenue is negative in very heterogeneous places and where ethnic similarity is very low since illegibility can make tax administration very inefficient. In summary, I expect increases in fiscal capacity to be higher in more homogeneous administrative units, independent of the level of similarity. I also expect increases in fiscal capacity to be higher in more similar administrative units, independent of the level of homogeneity.[10] These hypotheses follow:

---

[10]One worry can be that these two measures are highly correlated with each other. Even though it is theoretically likely, this is not the case in the dataset I use, as I discuss in Section 3.2.3.

Hypothesis 1a: *Wartime changes in the levels of fiscal capacity are higher in more homogeneous administrative units.*

Hypothesis 1b: *Wartime changes in the levels of fiscal capacity are higher in more similar administrative units.*

The second set of hypotheses I formulate are about the costs of state building. It is critical to test the argument's implications regarding diversity increasing these costs as it is an important link in the chain from diversity to changes in fiscal capacity. To reemphasize, I expect collecting taxes to be more expensive in more illegible places. More specifically, the state must invest more in more diverse provinces to be able to reap a given amount of revenue. Thus follow these two hypotheses:

Hypothesis 2a: *The costs of investment in fiscal capacity are lower in more homogeneous administrative units.*

Hypothesis 2b: *The costs of investment in fiscal capacity are lower in more similar administrative units.*

2.4   Scope Conditions

The theory I presented above is likely to be bound by certain scope conditions. The first scope condition is an agriculture-heavy economy. An agricultural economy is more likely to be affected by the problem of illegibility because agricultural production is often scattered over a much larger and more difficult to reach area. Agricultural production is also more likely to be observed and taxed by state agents because in order to assess the output and taxes on the agricultural products, the state agents need to be there right after the products are harvested. Otherwise, it will be impossible to prove how much the agricultural producer had produced. Such problems can exist in taxation of other forms of economic activities such as industrial production or trade, and taxation of assets

such as property. However, I do not expect the problems of illegibility to be as salient as it is for agribulture for these other modes of economic activity.

The second scope condition is indirect rule. One critical aspect of the theory I presented above is the bargaining between the state and the local intermediaries during state building. The problem of illegibility naturally should cause similar problems in contexts with direct rule, where a state that cannot learn much about the population will face challenges in ruling and taxing these populations. However, these problems should be more evident under indirect rule where the state needs the intermediaries to rule, and has to sideline them when trying to centralize its rule and tax collection.

Chapter 3: Research Design and Empirical Strategy

Based on the theory described in Chapter 2, this chapter presents the empirical strategy and the data to be used to evaluate the theory. In order to test the hypotheses, and also further empirical implications of the mechanism that I suggest that drive the relationship between diversity and fiscal capacity building outcomes, I use several quantitative datasets from archival and historical sources, some of which are original, and some of which are from secondary sources. In addition to statistical analyses of these quantitative data, I also qualitatively analyze correspondences such as letters, telegrams, and reports written by Ottoman administrators in order to provide additional support for my arguments. In the following sections of this chapter, I first present a very simple summary of my empirical predictions and then describe the empirical strategy, including case selection, the data I use, and the variables in more detail.

As I outlined before, in the empirical chapters, I mostly focus on ethnolinguistic diversity, and consider religious diversity only as robustness checks of the main quantitative empirical analyses in Chapter 6. I also discuss evidence relating to religious diversity in the qualitative analysis sections.

3.1 Hypotheses and Expectations

In order to provide an overall picture of how the following empirical chapters are organized, in Table 3.1, I present a summary of what each empirical chapter focuses on. Table 3.1 presents the hypotheses or expectations that each chapter aims to test and the empirical evidence that I will use in each chapter to test these hypotheses and expectations. A more detailed discussion of the empirical strategy will follow in the sections, or in the empirical chapters below.

Table 3.1: Hypotheses, Expectations, and Empirical Evidence Provided in the Following Empirical Chapters

Chapter	Hypothesis or Expectation	Empirical Evidence
Chapter 4	Diversity undermines legibility	Census completion; correspondences
Chapter 5	Diversity undermines bureaucratic capacity	Bureaucratic assignment patterns; correspondences
Chapter 6	*Hypotheses 1a and 1b:* Wartime increases in fiscal capacity are lower in more diverse places	Local-level fiscal revenues; correspondences
Chapter 7	*Hypotheses 2a and 2b:* Diversity increases the costs of investment in fiscal capacity	Local-level expenses; bureaucrat salaries; correspondences

3.2 Empirical Strategy

3.2.1 Case Selection

In this study, I adopted a within-country design that engages in a comparison of subnational cases (Snyder 2001). I argue that a within-country comparison is preferable over a cross-country comparison for a study of fiscal capacity building like this for at least two of reasons. First, there can be subnational unevenness in that can be masked by national averages. In many contexts the very same state can have different levels of control, capacity, and tax revenues in different areas under its rule (Bersch, Praça, and Taylor 2017; Blaydes 2018; Garfias and Sellars 2021).

Second, fiscal capacity is often measured using the tax revenues of the central state. Such measures include tax revenues per capita and tax revenues as a share of GDP. One factor that renders this measure problematic in cross-national studies is that it is impossible to know that it is low fiscal capacity that keeps tax revenues low. For instance, another reason why tax revenues are low can be the government's unwillingness to tax or because the state does not need any funds. Unfortunately, however, there is no better alternative measure of fiscal capacity, at least to my knowledge. By focusing on a single country, I can use the same measure of fiscal capacity without worrying about such a problem because in a single-country design the state's willingness to tax will be held constant.

The empirical focus of the study is the Ottoman Empire in the late-nineteenth and early twentieth century. The Ottoman Empire is an ideal case for a study that investigates how diversity shapes state capacity with the wide within-country variation in the levels of diversity as well as levels of state capacity.

Most of the data I use in this study is from the second part of the nineteenth century and the first decade of the twentieth century, although I also discuss some evidence from other periods of the empire when it is appropriate. There are several reasons why I choose to study the period. The first and the most obvious one is data availability. The local-level fiscal revenue data I was able to collect covers the decades starting from the 1860s. A second reason that the analysis starts at this date is because the empire was organized into Provinces *(Vilayet)* as the highest level administrative units starting with the first Provincial Law *(Vilayet Nizamnamesi)* in 1864, and then later with the second Provincial Law in 1871. The end period for the analyses I adopted is the year 1910. After this year, the Ottoman Empire engaged in a constant warfare where it lost significant amount of territories. These wars are the 1911-2 war with Italy, the Balkan Wars of 1912-3, and World War I. Extending the analysis after this period would not have allowed comparison as the Empire lost a vast majority of its territories during these wars.

Some other reasons why I focus on this period are related to the scope conditions of the argument. The Ottoman Empire during this period had a mostly agricultural economy, and heavily relied on indirect rule, especially in tax collection and administration. Many taxes were farmed. As I have discussed in Chapter 2, I expect this theory to be more likely to hold for mostly agricultural contexts, because legibility is more critical for agricultural taxation compared to forms of urban or property taxation. The Ottoman empire was a mostly agricultural empire, and one major way how it financed the enlargement of its bureaucratic and military apparatus after the beginning of the Tanzimat Era (1839) was by putting the tax burden mostly on the agrarian taxpayers (Quataert 2005, p.71).

Another reason relating to the scope conditions is that the Ottoman Empire relied on indirect rule to rule and tax many areas under its control. The power struggles between the local inter-

mediaries and the center had long hindered fiscal capacity building attempts before the nineteenth century (Karaman and Pamuk 2010). Local intermediaries often were tax farmers who also had political offices and resisted the central state's centralization attempts; unwilling to lose their tax farming or other privileges. Even though with the efforts of the center from early- to mid-19th century, especially during the reign of Mahmut II (r. 1808-1839), the power of the big local intermediaries had significantly eroded, the challenge from intermediaries had not disappeared, but only changed its shape.

At the beginning of the Tanzimat era, large-scale local intermediaries had been eliminated and most tax farms were owned by smaller intermediaries. The rulers had also managed to impose rules that kept the tax farms at the small scale, for example limiting the size of the farms (such as to the village) and not allowing the same person to buy contracts in multiple districts (Özbek 2010, 2018). Yet, several attempts to abolish tax farming and centralize tax collection failed. One large-scale attempt to abolish tax farming and appoint officials to administer tax collection between 1839 and 1842 failed and tax farming was reinstituted (Pamuk 2016). Throughout the nineteenth century, there were other instances when the state centralized at least portions of agricultural tithe collection in certain regions. This was the case during 1853-1854, during the Crimean war, and also to some extent after 1877, the Russo-Ottoman war of 1877-78 (Özbek 2018).

The fact that during the Tanzimat era the local intermadiaries who had tax farms were more often smaller intermediaries, who were from the local populations makes this context a more appropriate one to evaluate the theory here. In the pre-Tanzimat era there was a 'grand coalition' (*büyük koalisyon*) of Istanbul-based elites and elites in the provinces, including different groups such as "innumberable contractors, agents, financiers, accountants and managers" (Pamuk 2004, p.17), which laid claim on most tax farming contracts. These contracts were often bought at the center by powerful actors and then subcontracted to their local partners at the periphery.

The part of my argument about diversity increasing the bargaining power of local intermediaries vis-a-vis the state is less likely to apply for a context where tax farming contracts were owned by such a large and powerful coalition since their strong hold on the contracts did not solely stem

from their strong bargaining power as a result of the problem of diversity. Instead, it also stemmed from their strength due to prominent actors having a stake in these contracts. In such a context, diversity's effect on bargaining power might be overshadowed by how other factors shape the bargaining power between this coalition and the state. On the other hand, after this coalition was weaknened and the state managed to decentralize the allocation of tax farming contracts, my argument about how diversity shapes bargaining power is more likely to apply. With the weakening of this coalition, the state was finally successful in directly awarding the contracts to local actors, and the auctions were more often in the localities. The local actors who got these contracts were powerful in their area of influence, but probably did not possess the strong ties to the powerful actors at the center, because of which the balance of barganing power between them and the state more likely dependent on local factors such as the population's diversity.

Finally, I argue that focusing on this specific era in the Ottoman Empire allows us to focus on wartimes as periods when rulers invest in fiscal capacity building. The literature finds that without constraints on the ruler, economic and political development, including fiscal capacity development, is less likely (Dincecco 2011; Besley and Persson 2011). In the Ottoman case, the start of the Tanzimat era in 1839 can be taken as a turning point to transition to better constraints on the ruler and stronger rule of law. With the Gülhane Edict of 1839 the Ottoman Sultan abrogated his right to confiscate any property at will, transitioning to an era with stronger property rights (Findley 1980; Arslantaş 2017; Arslantaş, Pietri, and Vahabi 2020; Cansunar and Kuran 2019). Indeed, Magiya (2022) finds that due to the more frequent wealth-sheltering behavior of the wealthy before this period, the Ottoman state could not increase its fiscal revenues during wartime, but after the transition to a stronger rule of law regime wars corresponded to periods when the state was able to build stronger fiscal capacity. This is why, by focusing on this period, we can examine how fiscal capacity building outcomes are different across areas with different levels of diversity during wartime, when rulers invest in fiscal capacity building.

3.2.2 Data

I will present the quantitative datasets I use here only briefly below, and leave the more detailed description to each corresponding chapter where those data are used. Next, I will describe the archival documents I qualitatively analyze in each chapter.

Quantitative Data

In this study, I use several different datasets, some of which are original and some of which are from other sources. In the first empirical chapter of this study, Chapter 4 I use two sets of data on census completion at the province and sancak levels that I compiled from a report about the censuses that was written to the Sultan. In Chapter 5, I use a dataset of governor assignments in the late Ottoman Empire by Tezcür, Magiya, and Popescu (2021).

In Chapter 6, I present the main dataset of this study, the local-level fiscal revenue data of the Ottoman Empire between the years 1868 and 1910. Product of my eighteen-month archival work, and constructed from hundreds of archival and historical documents, the dataset provides information on fiscal revenues at the province level. It also has, albeit sparser, information on fiscal revenues at the sancak level. In the same chapter, I also use data on individual-level tax assessment in the central kaza of the Monastir province, based on the property censuses conducted in the 1840s.

In Chapter 7, I use two different sets of data. The first is local-level expenditures of the Ottoman State for the fiscal year 1909-1910, from the Fiscal Statistics *(İhsaiyat-ı Maliye)* Journal published by the Ottoman State. The second is data on Mutasarrıf salaries from the years 1872-1873, as provided by Bouquet (2007).

Archival Correspondences

I will complement the quantitative empirical analyses in each chapter with qualitative analysis of letters, telegrams, and reports written by Ottoman administrators. These are manuscripts that are

located in the Presidential State Archives *(Cumhurbaşkanlığı Devlet Arşivleri)* in İstanbul.[1] In Appendix A, I provide the pictures of three sample documents, and their transcriptions in modern Turkish.

An important note of caution for these sections containing data from these sources is that the analyses which use these data cannot systematically evaluate any of the hypotheses due to the 'data generating process' at work here. For example, one can expect to find documents where bureaucrats would be complaining about the reasons why tax collection in a certain place is undermined (e.g., due to the population's diversity); however, it is unlikely that we would see a letter that when such processes go smoothly, in non-diverse areas. I evaluate the hypotheses in a more systematic fashion with the quantitative datasets. In the chapters where I present evidence from the correspondences, it is mostly aiming to present additional evidence which aims to pinpoint instances where the mechanisms I suggest that drive these results indeed can be observed in these sources.

Another issue that needs to be kept in mind about the 'data generating process' regarding the evidence from the archival documents is that the documents in the Ottoman archives are searched by keyword and after searching for the documents with the relevant keywords I could reach these documents. This means that I reached these documents after a specific selection process based on the keywords I am already interested in and the existence of phenomena such as the difficulties in counting the populations and problems in the administration due to diversity in these documents led me to them.

3.2.3 Variables

Below I describe the dependent and the independent variables I use in the quantitative empirical analyses. Because I use a different dependent variable from a different dataset in each chapter, I provide only a less detailed description of these variables here and discuss them in more detail in each chapter. Because I use two main independent variables that measure diversity in most quantitative analyses each chapter, I discuss these variables measuring diversity in more detail

---

[1]In citing the sources in these archives, the old name of the archives, Başbakanlık Osmanlı Arşivi *(Prime Ministry's Ottoman Archives)*, is still used.

Table 3.2: Summary of Dependent Variables Used in Quantitative Analyses by Empirical Chapter

Chapter	Main Dependent Variables
Chapter 4	Census Quality and Completion
Chapter 5	Governor Assigned to Birth Region; Governor Assigned to Previously Served Region
Chapter 6	Per Capita Tax Revenues
Chapter 7	Expense-to-Revenue Ratio; Type of Expenditures; Bureaucrat Salaries

below, and leave the description of the other independent variables to the chapters where they are used.

Dependent Variables

In the first empirical chapter of this study, Chapter 4, I use a report presented to the Sultan about the 1881-1893 population census in the empire (BOA.Y.PRK.A, 78/8) to code the quality and completion of the census in different administrative units in the empire as a measure of the legibility of the population to the state and its agents. In order to show that diversity put constraints on government assignment patterns in a way that could constrain bureaucratic capacity, I use Tezcür, Magiya and Popescu's data to construct variables on which region a governor was assigned to. In Chapter 6, the dependent variable is fiscal revenues per capita at the province level. Finally, in Chapter 7 I use the local-level revenue dataset together with another dataset on local-level expenditures to construct two dependent variables. The first is Expense-to-Revenue Ratio, to measure how much the state has to invest in a province to be able to extract a unit revenue from this province. The second dependent variable I construct with this data is Type of Expenditure, which measures the proportion of long-term and short-term investments that the state makes. Using another dataset on Mutasarrıf salaries in different sancaks, I conduct another set of analysis where the dependent variable is the salary of a Mutasarrıf in a given sancak. Table 3.2 presents a summary of what dependent variable I use in each chapter.

Main Independent Variables

There are two independent variables measuring the two different configurations of diversity. In order to measure heterogeneity, I use Ethnolinguistic Fractionalization (ELF). Following Alesina et al. (2003), I calculate ELF using the Herfindahl-Hirschman Index:

$$ELF: 1 - \sum_{i=1}^{N} s_i^2$$

where *s* denotes the share of each ethnolinguistic group *i* within a province . Higher value of ELF indicates higher heterogeneity (higher diversity). This measure of heterogeneity can be interpreted as the probability that two randomly chosen individuals come from different groups (Easterly and Levine 1997).

To measure similarity, I use a measure I call Ethnic Similarity (ES). The ES measure is the proportion of the Turkish ethnic population in the province. Higher value of ES indicates higher similarity (lower diversity). Each of these measures theoretically range between 0 and 1.

In order to construct these measures I mainly relied on the Ottoman censuses. In addition to them, I relied on the estimates and corrections provided by secondary sources. One challenging factor in constructing these measures was that the Ottoman censuses do not report any information about ethnicity, but only religious denomination. It is relatively easier to identify ethnicity for non-Muslim populations because religious denomination indicates ethnicity for most major non-Muslim groups in the empire (such as Greeks, Armenians and Jews). There were only minor issues to take into account regarding the ethnicity of non-Muslims. One issue was in the Arab regions of the empire. Christian Arabs were reported according to their sect, for instance Arabs in the Greek Orthodox faith were reported as Greeks (Karpat 1985a). For this reason, in the Arab regions, I assumed that those reported to be of Greek Orthodox or Latin Catholic faith were Christian Arabs and therefore coded these groups as ethnic Arabs. Another issue, although their numbers were far smaller, was the Protestant groups in Middle and Eastern Anatolia. I assumed that the Protestants in Middle and Eastern Anatolia were Armenians since there was significant Protestant missionary

activity in Anatolia and many Armenians had converted to Protestantism throughout the nineteenth century (Arpee 1936).

It was more complicated when it came the ethnic populations of the Muslim groups. Since the censuses provided no clues about the ethnic identities of the Muslims, I had to turn to secondary sources. For the number of Kurds, Turks and Armenians in the six provinces in Eastern Turkey I used the population estimates in Marashlian (1991) and based on his estimates and the census data I extrapolated the size of each ethnic group.[2] For several of the Arab provinces of the empire, I assumed that all those reported as Muslims were Arabs. These provinces are Baghdad *(Bağdat)*, Basra, Beirut *(Beyrut)*, Hejaz *(Hicaz)*, Syria *(Suriye)*, Tripoli of Libya (Trablusgarp) and Yemen. Indeed, in a report to the Sultan, it is reported that all the Muslim tribes in the Arabian Peninsula speak Arabic (BOA.Y.PRK.AZJ, 44/68). For Mosul *(Musul)* and Aleppo *(Halep)*, I estimated the proportion of Turks, Kurds and Arabs based on the British population statistics (Zamir 1981). Since parts of Trabzon and Erzurum provinces were lost by the Ottomans to Russians after the war of 1877-8, I additionally used the official Russian population statistics of 1897 in estimating the pre-1877 ethnic composition of non-Muslim groups in these provinces. For Trabzon, another source to rely on was the 1927 and 1965 censuses conducted by the Turkish Republic in order to estimate the statistics for Muslim groups of different ethnic groups. Finally, for the ethnic composition in the Balkan provinces, I relied on the estimates provided by Belgian magazine Ons Volk Ontwaakt (1912) and by Antonean (1975).

One possible worry about using the ELF and ES (Percent Turkish) variables together can be that they may be correlated, as the share of certain groups and measures of heterogeneity can be correlated (Kustov and Pardelli 2018). In this study, they are not highly correlated in a way that can cause a problem in the empirical analyses. As I show in Figure 3.1, their linear relationship is a very weak one.

---

[2]These provinces are also known as Vilayat-ı Sitte (literally, six provinces). They had significant Armenian populations and were political hotspots. They are Bitlis, Erzurum, Diyarbekir, Mamuretülaziz, Sivas and Van.

Figure 3.1: Scatterplot of Ethnolinguistic Fractionalization and Ethnolinguistic Similarity
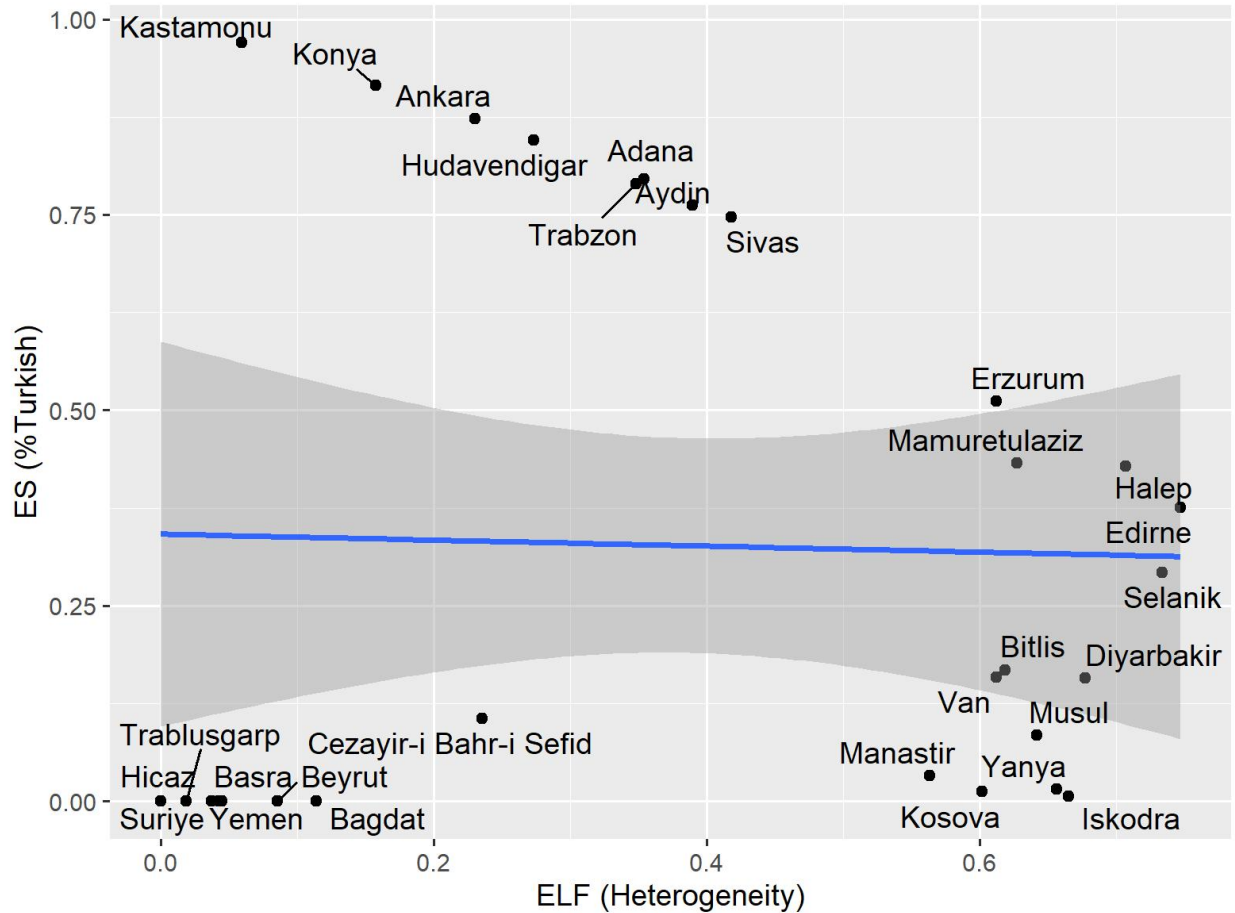
Alternative Independent Variables

Even though my argument can apply both to ethnic and religious diversity, I noted that most of my empirical analyses would focus on ethnic diversity. Yet, in order to be able to replicate some key empirical tests using measures of religious diversity, I constructed two different measures that include religious diversity of the populations. The first such measure I construct is Religious Fractionalization (RF). This is constructed using the Herfindahl–Hirschman Index. The religious identity used here is simply the religious identity of the individual, corresponding more closely with the Millet System in the empire, with one exception, the sectarian differences among Muslim populations. Sectarian differences are argued to have contributed to the illegibility of the populations to the state (Blaydes 2018). I build this measure mostly based on the Ottoman censuses,

which report religious affiliation. An extra source I utilized for the Sunni and Shia populations in the Baghdad, Basra and Mosul provinces is the British censuses, as reported in Zamir (1981).

I also constructed another measure of diversity according to the ethnoreligious identities of the population, again using the Herfindahl-Hirschman Index. This measure aims to include the diversity in terms of both to the ethnic and religious identities of the population. Within each ethnic group, each religious group is considered to be a separate group. Constructing it from the same sources that I used for constructing the ELF measure, I call it Ethnoreligious Fractionalization (ERF).

Chapter 4: Diversity Undermining Legibility: Evidence from Ottoman Censuses

In Chapter 2 I argued that informational capacity (i.e. legibility of the population) and bureaucratic capacity of the state are two necessary factors for the state to be able to build fiscal capacity. I also argued that diversity undermines each of these, legibility and bureaucratic capacity, and therefore constrains the state's fiscal capacity. In this chapter I provide evidence for the first of these arguments, by showing that the more diverse the population is, the more illegible it is to the state and its agents.

If a reliable census is available, any agent of the state, whatever their job is, can more easily access standardized information about the population. Such information is often critical for the state's control and rule over the populations, and is necessary to be able to conscript soldiers into standing armies as well as efficient tax collection. Anyone who glances at census records can learn how many people live in each locality, their demographic information, and the economic activities these people are engaged in. Therefore, censuses are important indications of the population's legibility to the state (Scott 1998; Lee and Zhang 2017; Brambor et al. 2020).

Ruling over various ethnic and religious groups, the Ottoman Empire had trouble collecting standardized and high-quality information about some populations under its rule. To put it differently, these populations were illegible to the Ottoman state. Diversity makes administering censuses more difficult because the dissimilar ethnic and religious identities and heterogeneity increase the costs of administering censuses and therefore the costs of obtaining information. For an outsider who does not speak the language, unfamiliar with local customs and lacking the necessary connections, counting the population and obtaining the relevant information is more difficult

compared to an insider who can speak the language, who is familiar with the local customs, and who may possess critical networks among the population.

One indication of the overall illegibility of the population in the Ottoman Empire is observable in the frequency and the quality of the censuses that the empire attempted to conduct over many decades. While the empire conducted the first census as early as 1831, this was not really a census in the modern sense since it only counted the households, but not individuals and covered only some parts of the empire (Karal 1943, cf. Karpat 1978). Later censuses starting from the late-nineteenth century offered the center higher quality data on many regions of the empire; however, even well into the twentieth century, Ottoman censuses were very unreliable in diverse regions where the state's power was very weak. The local populations could resist attempts to be counted (Yosmaoglu 2006) and the reports of local administrators to the center complain how difficult it was to count minority populations (BOA.A.MKT, 66/2). In fact, the empire never attempted to systematically count the populations in Hejaz and Yemen provinces in a census. In these provinces, the state's control was very weak, almost no Turkish-speaking population lived and the necessity of good command of Arabic language and knowledge of Arab customs to be an administrator there was explicitly acknowledged (BOA.DH.MKT, 1539/120). It also did not attempt to count the populations in the provinces of Baghdad, Basra and Mosul in the 1914 census, the last ever census by the empire. Populations in these provinces are also very diverse, consisting mostly of non-Turkish populations.

If diversity is at the root of the problem of illegibility, as I argue, then the state should be less likely to be successful in counting the populations, and therefore conducting censuses in more diverse places. In the following sections, I first demonstrate that the state was less successful in successfully conducting censuses in more diverse administrative units by using data from Ottoman censuses. Next, I provide further evidence from archival documents and secondary sources that again confirm the expectation that the diversity of the population undermined the ability of the state to extract information from the populations and therefore the legibility of the population.

4.1 Quantitative Evidence from Ottoman Censuses

4.1.1   Data and Variables

To evaluate my expectation that diversity should undermine successful censuses, I rely on the outcome of the 1881-1893 population census in the empire. From a report presenting the results of the census to the Sultan, it can be inferred in which first or second level administrative units the state had been able to complete the census (BOA.Y.PRK.A, 78/8). This is the outcome I am interested in. I measure this outcome both at the province (the first level administrative unit), and at the sancak level (the second level administrative unit). I conduct analyses at each level because for the former there are very few observations (twenty-nine observations) and low statistical power. In addition to this, measuring this outcome at the higher level can be misleading since within-province variation in census completion is lost when the province is considered as a whole. The problem with the sancak-level analysis is that the independent variable is missing for some units at this level and the analysis can be conducted only for a subset of the units.

Based on the census reports, I coded the census quality as an ordinal variable for the province-level data, lowest quality being 0, intermediate quality being 1, and highest quality being 2. For the analysis at the lower level administrative unit, sancak, the outcome variable indicates whether the whether the census was completed in a given administrative unit. I construct this outcome variable as a dummy, equaling 1 if the census was completed and 0 if not.

For each analysis, I exclude the administrative units where the census had never started, as it is impossible to know whether they would be completed or not. Table 4.1 lists how the outcome variable is coded for each province, and Table 4.2 lists how the outcome variable is coded for each sancak.[1]

For independent variables, I use the two measures of diversity that I described in more detail in Chapter 3. As a reminder, these are Ethnolinguistic Fractionalization (ELF) to measure hetero-

---

[1]The list on Table 4.2 does not include the sancaks where it was not possible to reconstruct ethnolinguistic distribution of the population.

Table 4.1: List of Provinces and the Census Quality at the end of the Census of 1881-93

Province	Census Quality (From 0 to 2)
Adana	Higher (2)
Aleppo (Halep)	Intermediate (1)
Ankara	Higher (2)
Aydın	Higher (2)
Baghdad (Bağdat)	Lower (0)
Basra	Lower (0)
Beirut (Beyrut)	Higher (2)
Bitlis	Lower (0)
Crete (Girit)	Did not start
Diyarbekir	Higher (2)
Edirne	Intermediate (1)
Erzurum	Intermediate (1)
Hejaz (Hicaz)	Did not start
Hüdavendigar	Higher (2)
Ioannina (Yanya)	Higher (2)
Kastamonu	Higher (2)
Konya	Higher (2)
Kosovo (Kosova)	Intermediate (1)
Mamüretülaziz	Intermediate (1)
Mediterranean Islands (Cezayir-i Bahr-i Sefid)	Intermediate (1)
Monastir (Manastır)	Lower (0)
Mosul (Musul)	Lower (0)
Shkoder (İşkodra)	Lower (0)
Sivas	Higher (2)
Syria (Suriye)	Intermediate (1)
Thessaloniki (Selanik)	Higher (2)
Trabzon	Higher (2)
Tripoli of Libya (Trablusgarp)	Did not start
Van	Lower (0)
Yemen	Did not start

geneity, and Ethnic Similarity (Percent Turkish) as a measure of similarity to the core/dominant group.

4.1.2 Results

The results are overall in line with the expectations that diversity undermines conducting censuses, rendering a population less legible to the state and its agents, as the censuses are more likely to be

Table 4.2: List of Sancaks and the Census Completion Status at the end of the Census of 1881-93

Province	Sancak	Census Status	Province	Sancak	Census Status
Adana	Adana	Complete	Hejaz	Mecca	Did not start
Adana	Mersin	Complete	Hejaz	Madina	Did not start
Adana	Cebel-i Bereket	Complete	Hejaz	Jeddah	Did not start
Adana	Kozan	Complete	Hüdavendigar	Bursa	Complete
Adana	İçel	Complete	Hüdavendigar	Ertuğrul	Complete
Aleppo	Aleppo	Incomplete	Hüdavendigar	Karahisar	Complete
Aleppo	Urfa	Incomplete	Hüdavendigar	Karesi	Complete
Aleppo	Maraş	Incomplete	Kastamonu	Kastamonu	Complete
Ankara	Ankara	Complete	Kastamonu	Bolu	Complete
Ankara	Yozgat	Complete	Kastamonu	Kangırı	Complete
Ankara	Kayseri	Complete	Kastamonu	Sinop	Complete
Ankara	Kırşehir	Complete	Konya	Konya	Complete
Ankara	Çorum	Complete	Konya	Niğde	Complete
Aydın	İzmir	Complete	Konya	Burdur	Complete
Aydın	Saruhan	Complete	Konya	Hamid	Complete
Aydın	Aydın	Complete	Konya	Teke	Complete
Aydın	Denizli	Complete	Mamuretülaziz	Mamuretülaziz	Complete
Aydın	Menteşe	Complete	Mamuretülaziz	Malatya	Complete
Ankara	Yozgat	Complete	Mamuretülaziz	Dersim	Incomplete
Ankara	Kayseri	Complete	Med. Islands	Rhodes	Incomplete
Ankara	Kırşehir	Complete	Med. Islands	Lesbos	Incomplete
Ankara	Çorum	Complete	Med. Islands	Chios	Incomplete
Baghdad	Baghdad	Complete	Med. Islands	Lemnos	Incomplete
Baghdad	Diwaniyah	Incomplete	Sivas	Sivas	Complete
Baghdad	Karbala	Incomplete	Sivas	Amasya	Complete
Basra	Basra	Incomplete	Sivas	Karahisar-ı Şarki	Complete
Basra	Muntafiq	Incomplete	Sivas	Tokat	Complete
Basra	Ammara	Incomplete	Syria	Damascus	Complete
Basra	Najd	Did not start	Syria	Hama	Complete
Beirut	Beirut	Complete	Syria	Hawran	Incomplete
Beirut	Acre	Complete	Syria	Kerak	Incomplete
Beirut	Tripoli	Complete	Trabzon	Trabzon	Complete
Beirut	Latakia	Complete	Trabzon	Canik	Complete
Beirut	Balqa	Complete	Trabzon	Lazistan	Complete
Edirne	Kırkkilise	Complete	Trabzon	Gümüşhane	Complete
Edirne	Gelibolu	Complete	Tripoli (of Libya)	Tripoli	Did not start
Edirne	Tekfurdağı	Complete	Tripoli (of Libya)	Homs	Did not start
Erzurum	Erzurum	Incomplete	Tripoli (of Libya)	Cebel-i Garbi	Did not start
Erzurum	Erzincan	Incomplete	Tripoli (of Libya)	Fezan	Did not start
Erzurum	Bayazıt	Incomplete	Van	Van	Complete
			Van	Hakkari	Incomplete

Table 4.3: Analysis of the Relationship Between Ethnic Composition and Census Quality at the Province Level

	Dependent variable:	
	Census Quality (0: Lowest, 2: Highest)	
	(1)	(2)
Ethnolinguistic Fractionalization (ELF)	0.621	0.436
	t = −0.993	t = −1.250
Percent Turkish	6.835	8.831
	t = 2.904***	t = 2.005**
Population (1000)		0.999
		t = −0.116
Sea Opening Dummy		1.765
		t = 0.427
Railroad Dummy		6.049
		t = 1.068
Average Elevation (km)		0.860
		t = −0.117
Land Border Dummy		0.393
		t = −0.939
Observations	26	26

Notes: Ordinal logistic regression. Odds ratios and t-values reported. *p<0.1; **p<0.05; ***p<0.01.

completed or to be of higher quality in administrative units with higher percentage of Turks and in more homogeneous administrative units.

For the province-level analysis I use an ordinal logistic regression model as the dependent variable is ordinal. With four provinces excluded from the analysis, there are only 26 observations and we have very low statistical power. Nonetheless, there is some support for the expectation that diversity makes it more difficult to count populations, as I report in Table 4.3.

In Model 1, I conduct the ordinal regression analysis without any covariates. Each diversity measure is standardized around their means. The odds ratio of the ELF variable is 0.621, indicating more homogeneous administrative units have lower quality censuses, although it is not estimated to be statistically significant. The odds ratio for the Percent Turkish variable is 6.685 and is signif-

Figure 4.1: Predicted Probabilties of Logistic Regression Models Predicting Census Quality at the Province Level by ELF Score
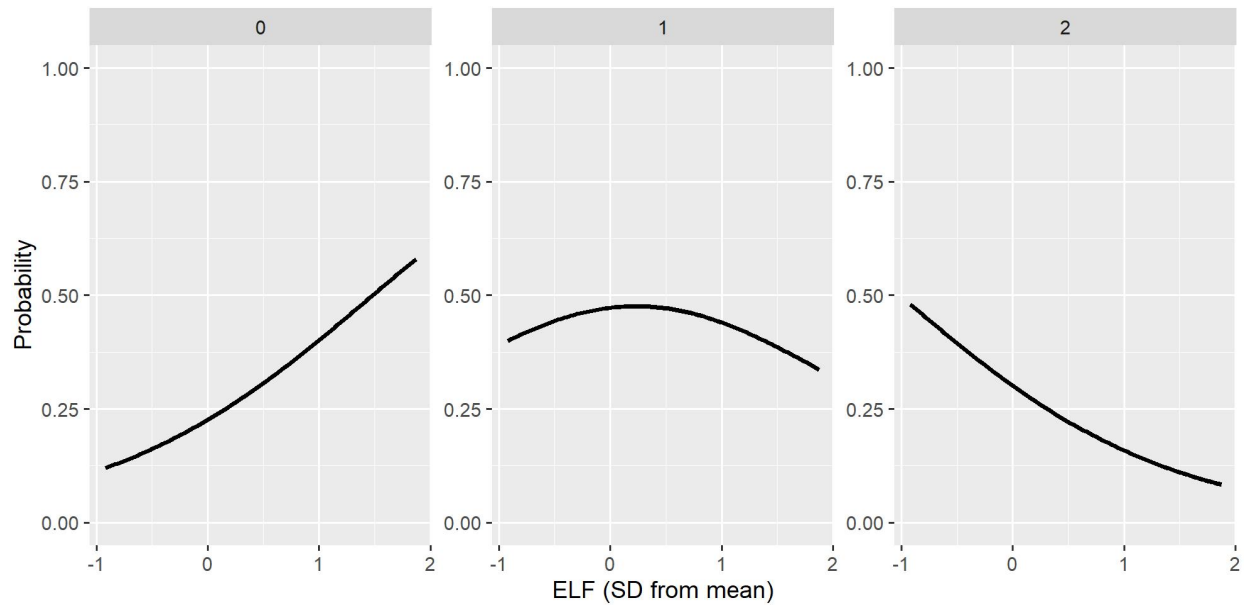
Notes: The figures present the predicted probabilities of census quality score by ELF score for a province, based on the ordinal logistic regression model used is Model 2 in Table 4.3, which controls for Percentage of Turks, Population size, Elevation, Land Border, Sea Opening and presence of Railroads. The plot on the left presents the predicted probabilities that the outcome variable is 0, the plot in the middle for the outcome variable being 1, and the plot on the right for the outcome variable being 2. All the covariates, including Percent Turkish, are kept at their median value.

icantly estimated, which suggests that for one standard increase in this variable, the census quality is more than 6.6 times more likely to be one level higher.

In Model 2, I include five covariates in the model, province's population (in 1000), whether the province has any sea opening, whether there are any railroads in the province, the average elevation of the province, and whether the province has any land borders to another state.[2]

The results in Model 2 are similar to those in Model 1. The odds ratio for the ELF variable declines to 0.436, but is still insignificantly estimated. This, again, is likely due to the very low number of observations in the analysis and low statistical power. The odds ratio of 0.436 indicates that for a one standard deviation increase in the ELF measure, the likelihood that census quality is

[2]Other covariates such as slope, area or distance from the capital city could have been added to the model. Yet, they were highly correlated with the main explanatory variables or with other covariates. As a rule of thumb, I excluded those covariates that had over a 0.5 correlation coefficient with the other variables.

Figure 4.2: Predicted Probabilties of Logistic Regression Models Predicting Census Quality at the Province Level by Percent Turkish Score
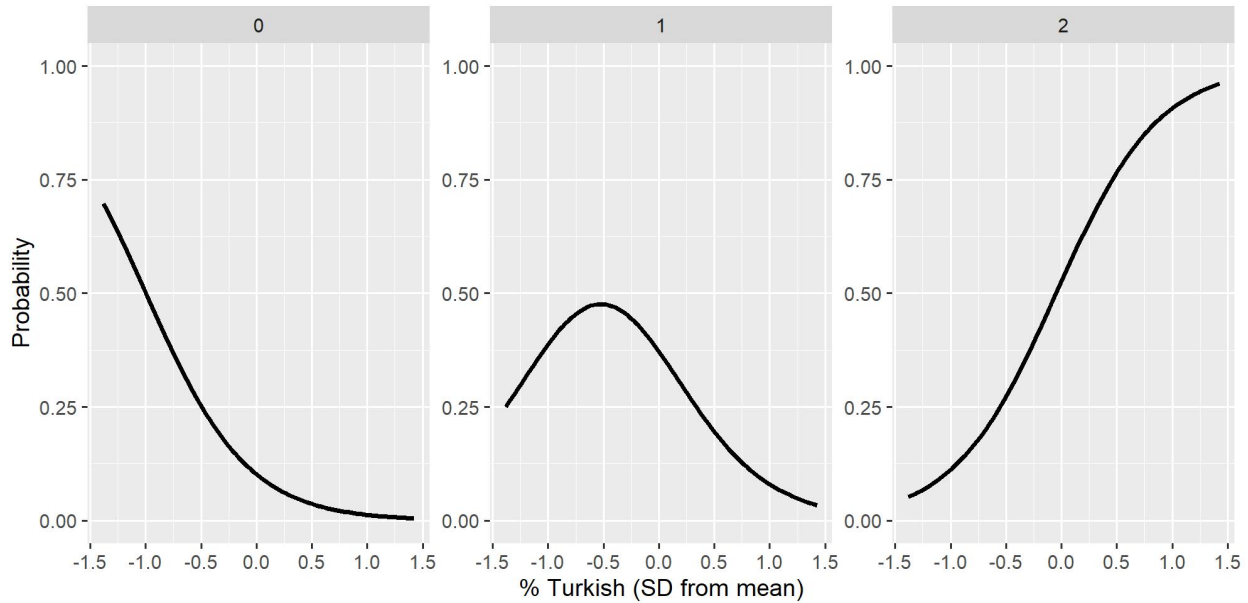
Notes: The figures present the predicted probabilities of census quality score by Percent Turkish score for a province, based on the ordinal logistic regression model used is Model 2 in Table 4.3, which controls for Fractionalization, Population size, Elevation, Land Border, Sea Opening and presence of Railroads. The plot on the left presents the predicted probabilities that the outcome variable is 0, the plot in the middle for the outcome variable being 1, and the plot on the right for the outcome variable being 2. All the covariates, including ELF, are kept at their median value.

one category more than doubles. This is substantively large and meaningful. The odds ratio for the Percent Turkish variable increases to 8.831, still estimated to be statistically significant.

Figure 4.1 presents the predicted probabilities for census quality scores according to the ELF measure, and Figure 4.2 presents the predicted probabilities according to the Percent Turkish measure, according to Model 2 in Table 4.3.

I report the results of the logistic regression analysis for the sancak-level data in Table 4.4. As I mentioned above, I had to exclude some of the sancaks of the empire in the Balkans and Eastern Anatolia from the analysis since sancak-level ethnicity data for the Muslims in these units are impossible to reconstruct. I exclude all the sancaks of Kosovo, Manastır, Mosul, Thessaloniki, Ioannina, Diyarbakir and Bitlis provinces. I was able to reconstruct the ethnicity data for three sancaks of Edirne province (Tekfurdağı, Kırkkilise and Gelibolu) from the censuses of the Turkish Republic but had to exclude the remaining two sancaks of this province (Gümülcine and Dedeağaç)

Table 4.4: Analysis of the Relationship Between Ethnic Composition and Census Quality at the Sancak Level

	Dependent variable:	
	Census Completed Dummy	
	(1)	(2)
Ethnolinguistic Fractionalization (ELF)	0.448	0.331
	t = −2.044**	t = −2.034**
Percent Turkish	9.959	7.171
	t = 3.407***	t = 2.383**
Average Slope		1.048
		t = 0.312
Land Border Dummy		0.315
		t = −0.949
Sea Opening Dummy		7.126
		t = 1.659*
Area (km sq.)		1.000
		t = −0.772
Observations	70	64
Log Likelihood	−25.248	−17.424
Akaike Inf. Crit.	56.496	48.849

Notes: Ordinal logistic regression. Odds ratios and t-values reported. *p<0.1; **p<0.05; ***p<0.01.

that was not included within the borders of the Turkish Republic and therefore where Turkish censuses are not available.

Having a higher number of observations here, the coefficients for each of the diversity measures are in the expected direction and are estimated to be statistically significant in each of the models in Table 4.4. This indicates that the results are in line with my expectations. Censuses are significantly less likely to be completed in sancaks with higher ELF measures, and significantly more likely to be completed in sancaks with higher percentage of Turkish population. Model 1 in this table reports the results without any covariates and Model 2 reports the results with covariates.[3]

The odds ratio for the ELF variable is 0.448 in Model 1 and declines to 0.331 in Model 2. These results indicate that one standard deviation increase in the ELF measure of the sancak makes it twice more likely that a census is not completed according to Model 1, and three times more likely according to Model 2. The coefficient of the Percent Turkish variable being 9.959 in Moel

---

[3]I again excluded covariates that had high correlation with the diversity measures, or with each other. In addition to this, the Railroad Dummy variable does not exist at this level.

Figure 4.3: Predicted Probabilties of Logistic Regression Models Predicting Census Completion at the Sancak Level
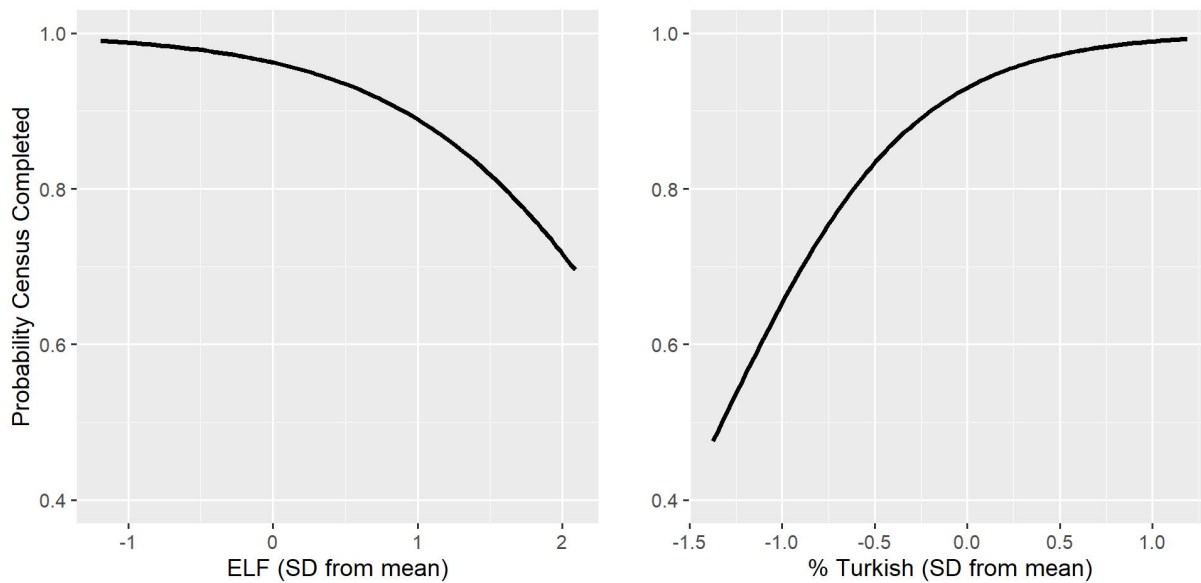
Notes: The figure on the left presents the predicted probabilities of completing the census by ELF scores of the sancaks and the figure on the right presents the same by the ES scores (% Turkish population). The logistic regression model used is Model 2 in Table 4.4, which controls for Slope, Land Border, Sea Opening and Area. All covariates are kept at their median value.

1 and 7.171 in Model 2 means that one standard deviation increase in the percentage of Turks in a sancak makes the census almost 10 times and over 7 times more likely to be completed, respectively. Figure 4.3 presents the predicted probabilities of census completion by ELF and ES scores according to the results in Model 2 of Table 4.4.

4.2 Evidence from Archival Documents and Secondary Literature

In the previous section I demonstrated that the diversity of the population undermined the censuses by the Ottoman state. In order to corroborate these findings, in this section I go on to present qualitative evidence from archival documents as well as secondary sources that present further evidence that it was more difficult for the Ottoman state to obtain information about and count populations that were not Turkish as well as those populations that were more heterogeneous.

For many states, acquiring knowledge about their peripheries has always been a challenging issue. Possessing little information about the periphery, the state agents from the center have often

perceived the peripheral regions as backwards areas and adopted discourses that label the populations in these regions as 'savages', whether in France (Weber 1976), or the Ottoman Empire (Makdisi 2002; Deringil 2003). Not understanding the languages or the customs of these populations were important reasons that reinforced these images of peripheral populations as 'savages' (Weber 1976; Deringil 2003).

In general, it is difficult to obtain information from and about populations in rural contexts, where the population may not be willing to share information with state officials whom they see as strangers, and where peasants limit their interaction with non-peasants (Weber 1976). Linguistic and cultural differences makes it even more difficult for the state's agents to communicate with, and therefore acquire information about the peripheral and rural populations. In the Ottoman Balkans, the state had to rely on local clergy to be able to communicate with the populations in villages and towns because they spoke local languages as the state agents often could not speak these languages (Karpat 1982). In his book *Peasants Into Frenchmen*, Eugene Weber discusses several examples where officers or merely simple travelers ran into linguistic barriers in the linguistically diverse French countryside in the nineteenth century, and had difficult time acquiring information as a result (1976, pp.45-82).

The Ottoman state had very poor information about its peripheral regions and the populations inhabiting these regions (Reinkowski 2017). Means to acquire knowledge about the customs of the population such as ethnographies or travel books were limited (Makdisi 2000). On top of this, the state lacked the resources to acquire necessary knowledge on a regular basis (Anscombe 1997). In order to properly control and tax the population in an area, any state would need to have high quality and up-to-date information that records each of the state's settlements, their geographical locations and how to reach there. In a report *(layiha)* he wrote in the year 1909 about the province of Yemen, Muhammed Hilal Efendi, a bureaucrat serving in this province, complains that the names of the villages of this province are available neither in the local-level administrative centers, nor in the records of the central state because the state had never conducted censuses in

this province (BOA.Y.EE. 58/33). In other words, the state and its agents did not know even the names and locations of many settlements in this province.

The Arab regions of the empire, including the province of Yemen, had very low proportions of Turkish speakers and in these regions the state always had more difficulty in obtaining information about the population and controlling them compared to other regions under its control. The inability of the state to count the populations or its properties was nothing special to the census of 1881-93, which is the focus of the empirical analyses in the previous section, but also is evident in other population and property censuses. One example is the census that was conducted in the early 1830s, which did not include Arab regions, Albania, Bosnia and South-East Anatolia (S. J. Shaw and E. K. Shaw 1977). None of these regions had any sizable Turkish population and many of them were very heterogeneous. Another example is the property census of 1845. The state was able to conduct this census in most areas under Ottoman control, with the exception of the Arab regions (Kaya and Terzibaşoğlu 2009).

Another example of a report from the province of Yemen indicates how little the administrators serving in the region knew about the populations inhabiting this province (Y.PRK.AZJ. 44/68). This twelve-page report prepared by a bureaucrat serving in Yemen apparently aimed to share with the center some information about the population. However, there is little in this relatively long report that the center can use for the purposes of projecting its control in this area. After mentioning the names of the ethnic and religious groups that live here, and only vaguely describing which groups approximately live what part of the province for around a page, the author of the report switches to describing minute details about the religious beliefs and practices of some of the Muslim sects that inhabit the province of Yemen. Ideally, the state trying to obtain information about the characteristics of the population in a region would have received more detailed information about in which locality which specific groups lived and what kind of economic activity they were engaged in, so that it could use such information for the purposes of controlling and taxing these populations. I argue that this report failing to achieve mentioning such useful information

54

but instead spending time on details, that are of dubious usefulness, is an indication that the state's agents here possessed little useful knowledge about the populations.

There are many complaints from the Ottoman bureaucrats in the periphery who had trouble counting the minority populations. The examples include areas in today's Greece, where they report not being able to find the Christian Orthodox Aromanians *(Ulahs)* in their villages in 1886 during the census of 1881-93 in Servia *(Serfiçe)* (BOA.DH.MKT. 1373/121), and not being able to finish the census on time in Edessa *(Vodina)* (BOA.TFR.I.SL. 84/8314). The examples also include not being able to complete the census on time in mostly Armenian and Kurdish-populated Van in Southeastern Anatolia (BOA.ŞD. 1881/19), in Arab-populated Nazareth (*Nasıra*) in the Levant (BOA.DH.MKT. 1563/25), and the provinces of Mosul and Baghdad (BOA.DH.MKT. 1542/96). Another example where we can directly observe a complaint about diversity of the population is from the year 1847. Mehmed Kamil, müşir of Sidon *(Sayda)* in today's Lebanon, where most of the population were Arabs, writes a letter to the center where he describes the difficulties he has encountered or expects to encounter during the census he is supposed to conduct (BOA.A.}MKT. 66/2). Here, he puts a lot of emphasis on the fact that this census is necessary to start military conscription and direct tax collection in the area. While he reports that he expects the census to proceed smoothly in some areas, he discusses potential problems that can arise due to the existence of large Christian populations and also because the populations are very mixed in this region. He notes that the notables of the communities in these regions should be summoned and that they should be informed about the benefits of a census since otherwise the spreading rumors (about the census) among the Christian communities here could generate resistance.

Another example from another Arab region, Tripoli of Libya *(Trablusgarp)* in the year 1901 exemplifies another tricky aspect of counting diverse populations (BOA.DH.MKT. 2485/85). In a telegram written to the office of the prime minister, the officials in the region inform the prime minister that the officials and the records that they were expecting from the center for a population count to be conducted here have not been received yet. Emphasizing that further delay in the count would be undermining any benefits it could generate, the telegram explains the intention of the

officials here to start the count with the help of the military officials and the notables in the area, who would travel to different localities in order to help the population count. It is difficult to know whether the help of these groups would be necessary had the census officials and the records that were suppose to arrive from the center arrived on time. Yet, very often military officials and local notables were indispensible for populations counts and tax collection in diverse areas (L. Anderson 1986; Doumani 1994).

Having to rely on the local intermediaries when counting the population was not ideal for the state because they had incentives to undermine the censuses and undercount the population, and relying on the military was costly. During the population count of Nablus, the Ottomans had sent military commander Muhammed Pasha to the city of Nablus with large numbers of troops in order to help the population count in Nablus and its hinterland (Doumani 1994). The reason he was accompanied by troops was probably the need to intimidate the local intermediaries into cooperation with the state. In this very case, the local intermediaries used their relative informational advantage to restrike a bargain about the actual census count numbers as soon as Muhammed Pasha left. Furthermore, one year after Muhammed Pasha's arrival, they petitioned the government officials to decrease the official population count of the area by 5,000 to 29,563 people from the original count of 34,563 by alleging that there were mistakes in the original count in addition to mistakes when adding up the numbers. As a result, they asked their tax burden to be decreased by 100,000 piasters since each individual was responsible for 20 per head. Doumani goes on to note that he could not find any responses to these petitions, so that the outcome of this bargaining process is unknown. However, population counts are costly, and sending in the military again would make it even costlier. For this reason, he concludes that "Unless the central government was willing to pay and investigate and retake the census, it would have no option but to accept these assertions at face value" (1994, p.8).

The final example I will mention here is from the context of post-Ottoman Iraq under Saddam Hussein in the late-twentieth century. Blaydes argues that the Ba'athist regime of Saddam Hussein, which was dominated by Sunni Arabs, faced challenges in monitoring the activities of Kurdish

Iraqis due to linguistic differences, and of Shi'a Arabs due to the difficulty of penetrating their networks (2018). This reflects a similar pattern to what I have discussed above. Groups that speak a different language from the core/dominant group's language, or groups that have a different religious identity than the core/dominant group's religious identity are more difficult for the state's agents to penetrate and obtain information from.

4.3 Conclusion

Censuses are important indications that the state and its agents have rendered a population legible and that they have standardized information about the population which can be used to extend the control of the state over the society. Trying to extend its control over its populations, tax them efficiently and achieve universal military conscription, the Ottoman State attempted to systematically conduct population censuses starting in the nineteenth century.

In this chapter I presented evidence from Ottoman censuses in addition to evidence from archival and secondary sources which aimed to demonstrate that more diverse populations are more illegible to the state and its agents. Obviously, the census data I used is not very high quality and it may not be the perfect quantitative data to measure the legibility of the population. The province-level data has very few observations which limit statistical power, and aggregating census quality at the highest level may mask within-province variation in census quality. In order to alleviate this problem, I used census outcomes at the sancak-level, the second-level administrative unit. While the outcome measure is more reliable at this level, the analysis for sancaks suffer from lack of the ethnic demography data for some sancaks, and therefore the analysis contains only a subset of the population.

In order to have better measures of the legibility of the population, there are several paths forward for data collection. One can be to obtain census records containing individual-level data and compare age-heaping across different localities to see if age-heaping, a measure of illegibility (Lee and N. Zhang 2017) increases with diversity.

Another path forward can be to construct a local-level cadaster quality measure, following D'Arcy, Nistotskaya, and Olsson (2019) that aims to capture how detailed information the state had about its territories, as cadastral information, like surveys, can communicate standardized and objective information to outsiders (Scott 1998).

Chapter 5: Diversity Undermining Bureaucratic Capacity: Evidence from Governor Assignments

After the bloody civil conflict in Mount Lebanon and Damascus in 1860, the Ottoman state granted the heterogeneous Mount Lebanon area an autonomous status and put it under special administration due to the pressure of Western powers. Ussama Makdisi describes the ceremony on 18 July 1861, when the proclamation of this special administration, *Mount Lebanon Mutasarrifate*, was announced to the people of Mount Lebanon as thus: "Turkish-speaking Ottoman officials, flanked by French, British, Austrian, Prussian, and Russian representatives, had the proclamation read in Arabic to the assembled crowd of notables and esteemed "faces" of the society." (2000, p.160).

Such an interaction between Turkish-speaking Ottoman officials and the populations in the periphery of the empire, who spoke another language, was not a rare phenomenon. The frequent mismatch between the officials and locals languages and cultures was a critical factor in contributing to the empire's bureaucratic capacity being constrained. The language of the administration in the Ottoman Empire was Turkish, which meant that a majority of the Ottoman bureaucrats spoke this language at least to some extent (Bouquet 2007; Kırmızı 2007). However, unless they came from minority populations, were educated in minority languages, or served in certain minority regions where they invested in learning the minority languages in these areas and became acquainted with the customs of the populations here, they did not speak the languages spoken in and understood the culture of the regions where they served. Due to the diversity of the population in the Ottoman empire and the many different languages and customs among the populations, such a mismatch between the bureaucrat's language and the language spoken in the area occurred frequently and was an important problem for the state.

In Chapter 2, I argued that, in order to build fiscal capacity, a state needs to have built both informational capacity (or rendered the population legible), and also needs to have built bureaucratic capacity. In Chapter 4 I demonstrated that diversity of the population undermines its legibility to the state and its agents. In the following sections of this chapter, I will focus on the second part of this argument and demonstrate that diversity of the population can undermine the state's bureaucratic capacity by utilizing evidence from governor assignment patterns, and also corroborating and complementing this evidence with further evidence from archival documents and secondary sources.

5.1 Evidence from a Dataset of Governor Assignments

In this section, I utilize a dataset about governor assignment patterns in the Ottoman Empire (Tezcür, Magiya, and Popescu 2021) in order to test further empirical implications of my argument, that diversity undermines state's bureaucratic capacity. In order to do increase the ability of the bureaucrats to accomplish their intended actions, a bureaucratic structure would need more skilled and experienced bureaucrats.

An ideal quantitative test involving bureaucratic capacity in this case would be able to measure bureaucratic capacity at the local level and test whether diversity is associated with bureaucratic capacity in the expected direction. Unfortunately any data that could help measure bureaucratic capacity at the local level in the Ottoman Empire do not seem to be available. Instead, I demonstrate how diversity affects governor assignment patterns in ways that are likely to constrain bureaucratic capacity, as it restricts the potential appointments of more skilled and more experienced bureaucrats.

The difficulty of acquiring knowledge about the diverse peripheral populations often made the Ottoman center to appoint officials with some knowledge of, and experience in the same or neighboring regions, as historians have observed (Anscombe 1997). However, the number of such officials who could solve the informational problems more easily as they could speak local languages and were familiar with local customs were limited due to diversity.

This leads me to expect a potential bureaucratic appointee with origins from a certain region to be more likely to be assigned to a province in his region of origin where his knowledge of the language and the customs in the population should be more useful. In addition to this, a potential bureaucratic appointee who has served in a region with certain characteristics should be more likely to be assigned to a province with similar characteristics to the previous regions where he served and has acquired information about the population. However, I argue that each of these assignment patterns will be more likely to be observed across homogeneous provinces, which are more likely to resemble each other than heterogeneous provinces can resemble each other. Furthermore, the knowledge that a bureaucrat possesses can be useful in an area where the population uniformly possesses the characteristics for which this knowledge can be used (i.e., language or customs). However, such a knowledge will be less likely to be useful in more heterogeneous areas, where many other groups also reside and for whom these the knowledge of the language and customs will not be useful. Serving in a region where this knowledge can be used for only a limited proportion of the population does not pose as significant advantages as serving in a region where this knowledge can be used for a much larger proportion of the population.

Each of these restrictions on the possible posts that the bureaucrats can be assigned are likely to constrain bureaucratic capacity. I argue that if the administrators can only be assigned to a more limited subset of positions based on their knowledge, the state's capacity can suffer via two mechanisms. First, the bureaucrats who serve in a narrower geographic area for a long period of time can build patronage relations that can benefit themselves and harm the state. Indeed, for centuries, the Ottoman State suffered from agents who established powerful ties where they served and challenged the state's power (Barkey 1994). In order to prevent this, the Ottoman State periodically shuffled its agents (Barkey 1996). The second mechanism, a more direct one, occurs because a given public servant cannot be assigned to any position where they could otherwise serve the state better, but the areas where they can do their job efficiently is limited by the languages they speak and the networks they can establish due to their identities and familiarity with customs.

In the following empirical analyses, I first show that being from a region with homogeneous characteristics makes a potential governor appointee more likely to be appointed to a province in this region, while we do not see such a pattern for those potential appointees who are from heterogeneous regions. Second, I show that having previously served in a province with homogeneous characteristics makes a governor more likely to be reappointed to another homogeneous province within the same region, where the characteristics of the population are not very different, and that this likelihood of appointment increases with the homogeneity of the province to which a governor is assigned. Even though the quantitative analyses I conduct in this section are at the highest-level administrative units, province, since I do not have any data about bureaucratic assignments at the lower levels, these patterns also likely holds at the lower levels. I provide further evidence for these patterns at different levels of administrative units in Section 5.2 below.

5.1.1 Classification of Provinces

I classified different provinces into four different geographical categories (plus the capital region which is not listed) that include adjacent provinces that are similar to each other in certain characteristics. Among these four categories, two of them contain homogeneous provinces that are also very similar to each other in the ethnic characteristics of the population. The first of these is are what I label Anatolia, comprising of the Middle and Western parts of today's Turkey, which had a majority of Muslim Turkish speakers. The second is Arab provinces which had a vast majority of Arabic speaking populations. Picking any two provinces within either of these regions, the chances are that the two provinces looked very much like each other in terms of ethnic composition.

The other two categories consist of heterogeneous provinces. The Balkan provinces include provinces that were in the European territories of the empire and had very high percentage of Christian populations who belong to different ethnic groups and religious denominations. There were significant numbers of Orthodox, Catholic and Muslim Albanians, Muslim Turks, Orthodox Serbs, Muslim Bosniaks, Orthodox Bulgarians, Orthodox Vlachs, Orthodox Greeks, Muslim Pomaks, and Jews. In the provinces that I include under Vilayat-ı Sitte *(Six Provinces)*, in what is

62

Table 5.1: Classification of Provinces

Region	Identity	Provinces	Mean ELF
Anatolia	Homogeneous; Majority Turkish	Adana, Ankara, Aydın, Hüdavendigar, Kastamonu, Konya, Trabzon	0.26
Arab	Homogeneous; Majority Arab	Aleppo (Halep), Baghdad (Bağdat), Basra, Benghazi (Bingazi), Beirut (Beyrut), Hejaz (Hicaz), Mosul (Musul), Syria (Suriye), Tripoli of Libya (Trablusgarp)	0.18
Balkans	Heterogeneous; Mostly non-Turkish Christians	Bosnia (Bosna), Crete (Girit), Danube (Tuna), Edirne, Ioannina (Yanya), Kosovo (Kosova), Mediterranean Islands (Cezayir-i Bahr-i Sefid), Monastir (Manastır), Shkoder (İşkodra), Thessaloniki (Selanik)	0.58
Vilayat-ı Sitte (Six Provinces)	Heterogeneous; Significant Armenian and Kurdish Populations	Bitlis, Dersim, Diyarbekir, Erzurum, Hakkari, Mamüretülaziz, Sivas, Van	0.51

Note: Several of the provinces are short-lived, either being abolished and joining other provinces, being occupied, or being annexed by foreign powers. The reason why there are eight provinces listed under the category *Six Provinces* is because the province of Hakkari was created in 1881 and rejoined the province of Van in 1889, and the province of Dersim was created in 1883 before it was merged into the province of Mamüretülaziz in 1889 (Sezen 2017).

today's Eastern Turkey, there were large numbers of Armenian, Kurdish and Turkish populations in addition to many other less populous minorities. Picking any two provinces within either of these two regions, one would be less likely to see an ethnic composition that looked alike each other.

Table 5.1 presents the provinces that fall into each category and the mean ELF scores of the provinces within each region. There is clearly a stark difference across Anatolian and Arab provinces, with ELF scores 0.26 and 0.18, respectively, while the Balkan provinces and Vilayat-ı Sitte have ELF scores of 0.58 and 0.51.

Table 5.2: Frequency Distribution and Proportion of Governor Assignments by Governor Birth Region/Ethnicity

	Anatolia	Arab	Balkans	Caucasus	Kurdish	Non-Muslim	Tatar
Anatolia	100 (0.33)	2 (0.11)	47 (0.27)	12 (0.19)	8 (0.28)	0 (0.00)	1 (0.11)
Arab	85 (0.28)	8 (0.44)	44 (0.25)	18 (0.29)	8 (0.28)	0 (0.00)	6 (0.67)
Balkans	61 (0.20)	3 (0.17)	40 (0.23)	20 (0.32)	2 (0.07)	7 (1.00)	2 (0.22)
Capital	4 (0.01)	0 (0.00)	3 (0.02)	1 (0.02)	0 (0.00)	0 (0.00)	0 (0.00)
Vilayat-ı Sitte	54 (0.18)	5 (0.28)	39 (0.22)	11 (0.18)	11 (0.38)	0 (0.00)	0 (0.00)

Note: The rows indicate assigned province, and the columns indicate governor birth region or ethnicity. Column proportions reported in parantheses.

5.1.2 Data and Variables

In order to test the first expectation, that a potential bureaucratic appointee should be more likely to be assigned to his region of origin if this region consists of homogeneous provinces, I used the dataset of Tezcür, Magiya and Popescu (2021), which provides information on each provincial governor appointment and the biographical information of these governors in late Ottoman Empire (1875-1918). The dependent variable I use is the region of the assignment, and the independent variable is birth place/ethnicity.

For the second expectation, that within-region reappointments of governors should be more likely for regions with homogeneous provinces, I again use the data by Tezcür et al. (2021) and create a dummy variable that equals 1 if a governor assigned to a province has ever previously served as a governor in any of the provinces within the same region, and 0 if otherwise.[1] The independent variable for this analysis is the ELF score of the province to which the assignment takes place. The variable on the ELF scores of provinces come from the same sources I described in Chapter 3.

5.1.3 Results: Assignment Patterns According to Birth Region/Ethnicity

Table 5.2 presents the frequency distribution and proportions of the governors' assignments by their birth region/ethnicity. I will only discuss the assignment patterns for governors who were

---

[1]An ideal version of this variable would measure if a governor had served in this region in any capacity. However, this information is not available in the dataset.

Table 5.3: Frequency Distribution and Proportion of Governor Assignments by Governors Born in Anatolia

	Not Born in Anatolia	Born in Anatolia
Not Assigned to Anatolia	228 (0.77)	204 (0.67)
Assigned to Anatolia	70 (0.23)	100 (0.33)

Note: Column proportions reported in parantheses. $\chi^2 = 6.5583$, N = 602, $p = .01038$

born in Anatolia and those who were born in the Balkans because the other birth place/ethnicity categories contain very few observations. These descriptive patterns are in the expected direction. Those governors who were born in Anatolia and therefore were of Turkish ethnicity were more likely to serve in Anatolian provinces, while no such clear pattern exists for the governors from the Balkan provinces.

In order to observe these patterns more clearly, I recoded the birth region/ethnicity variables and also the assignment provinces into binary categories. Table 5.3 presents the frequencies and the proportions in the binary version for Anatolian provinces, and Table 5.4 presents it for the Balkan provinces. The patterns are clearer here. The difference that being from Anatolia makes with regards to being assigned to these provinces seem to be much larger than the difference that being from the Balkans makes with regards to being assigned to the Balkans. In fact, being from the Balkans seems to make little difference in these patterns. In order to test whether these relationships are statistically significant or not, I conducted chi-square tests.

A chi-square test of independence for the variables of being assigned to Anatolia and being from Anatolia indicates that the two variables are significantly related ($\chi^2 = 6.5583$, N = 602, $p = .01038$). For the relationship between being assigned to a Balkan province and being from the Balkans, the chi-square test fails to achieve statistical significance. ($\chi^2 = .067627$, N = 602, $p = .7948$). These results confirm the observations that being from an Anatolian province makes it more likely to be assigned to an Anatolian province, while there is no statistically significant change in this likelihood for a potential bureaucratic appointee who is from the Balkan provinces.

In order to check a more likely case within the Balkans where the relationship between birth place/ethnicity and assignment province, I conducted a similar analysis which considers whether

Table 5.4: Frequency Distribution and Proportion of Governor Assignments by Governors Born in the Balkans

	Not Born in the Balkans	Born in the Balkans
Not Assigned to the Balkans	334 (0.78)	133 (0.77)
Assigned to the Balkans	95 (0.22)	40 (0.23)

Note: Column proportions reported in parantheses. $\chi^2 = .067627$, N = 602, $p = .7948$.

Albanian governors were more likely to be assigned to Balkan provinces with substantial number of Albanian populations, which are still very heterogeneous. These provinces are Ioannina (Yanya), Kosovo (Kosova), Monastir (Manastır), and Shkoder (İşkodra).

The patterns in Table 5.5 reveal that in fact, being Albanian does not make it any likely for a governor to be assigned to any of the four provinces with significant Albanian populations. If anything, the relationship is in the opposite direction. Albanian governors are less likely than non-Albanians to be assigned to one of these four provinces, although the relationship is not statistically significant ($\chi^2 = .94224$, N = 602, $p = .3317$).

To recapitulate, these assignment patterns by birthplace or ethnicity suggest that if a governor comes from an area, thereby speaking the language of the population in a province that has a homogeneous population and is familiar with the customs of this province, his probability to be assigned to a province with similar characteristics is higher. The language skills and the knowledge of the customs can prove useful in any of homogeneous provinces that have populations that speak the same language and have the same culture. On the other hand, those who come from an area that is heterogeneous, where provinces are less likely to resemble each other and where the language they speak and the customs they are familiar with can only help them with a smaller proportion of the population of the province, are not more likely to be assigned to another province within the same region, probably because it did not present significant advantages compared to being assigned to any random province.

Table 5.5: Frequency Distribution and Proportion of Albanian Governor Assignments

	Not Albanian	Albanian
Not Assigned to "Albanian Provinces"	484 (0.89)	54 (0.93)
Assigned to "Albanian Provinces"	60 (0.11)	4 (0.07)

Note: Column proportions reported in parantheses. The four Albanian provinces have large Albanian populations, but are also home to many other groups and are very heterogeneous. These four provinces are: Ioannina (Yanya), Kosovo (Kosova), Monastir (Manastır), and Shkoder (İşkodra). χ^2 = .94224, N = 602, p = .3317

5.1.4 Results: Assignment Patterns According to Previously Served Places

In this section, I test the second expectation about bureaucratic assignment patterns I discusse above, that within-region reappointments of governors should be more likely in homogeneous regions. Table 5.6 reports the results of logistic regression analyses where the ELF score of the province predicts whether the governor who is assigned to a province has served in any province within the same region. In the first model, no covariates are included and in the second model covariates Population, Sea Opening Dummy, Railroad Dummy, Average Elevation, and Land Border Dummy are included as covariates. All of these are measured for the province to which the governor is assigned. Each model includes region fixed effects and standard errors are clustered at the province level.

The results confirm the expectation that diversity is less likely to allow within-region assignment of governments. As a province gets more heterogeneous, the probability that a governor is assigned within the same region declines. These findings are in line with the argument that knowledge acquired in a homogeneous province should be more likely to be useful in a province that is homogeneous with characteristics that are alike, but we should not see this pattern across heterogeneous provinces. Figure 5.1 presents the predicted probabilities that a governor assignment is within-province according to ELF scores.

Table 5.6: Analysis of the Relationship Between Heterogeneity and Governor Assignment Patterns

	Dependent variable:	
	Governor Assigned Within-Region Dummy	
	(1)	(2)
Ethnolinguistic Fractionalization (ELF)	0.710	0.738
	t = −3.744***	t = −2.559**
Population (1000)		1.000
		t = −0.412
Sea Opening Dummy		0.554
		t = −1.493
Railroad Dummy		1.034
		t = 0.123
Average Elevation (km)		1.000
		t = −0.388
Land Border Dummy		0.866
		t = −0.441
Region Fixed Effects	Yes	Yes
Observations	383	383

Notes: Logistic Regression. Odds Ratios with t-values in Parantheses. Standard errors clustered at the province level. *p<0.1; **p<0.05; ***p<0.01.

5.2 Evidence from Archival Documents and Secondary Literature

The secondary literature and the primary archival sources are replete with evidence where the language skills and local ties of bureaucrats or administrators were critical in extending the state's control and its ability to raise taxes in the periphery of the Ottoman Empire. The governors that the Ottoman state sent to the provinces often lacked sufficient knowledge about their region of assignment and this undermined their ability to rule effectively (Reinkowski 2017). In order to solve this problem the Ottoman state made bureaucrats start their careers in their birth regions as it was advantageous to have agents who was familiar with the language and customs of the population, and also who was preferably known by the population (Bouquet 2007).

The local notables had to use minority languages in their area of influence. For example, Fleming (1999) notes that Ali Paşa of Ioannina (also known as Tepedelenli Ali Paşa), who was

Figure 5.1: Predicted Probabilities that Governor is Assigned within-Province, 1876-1918
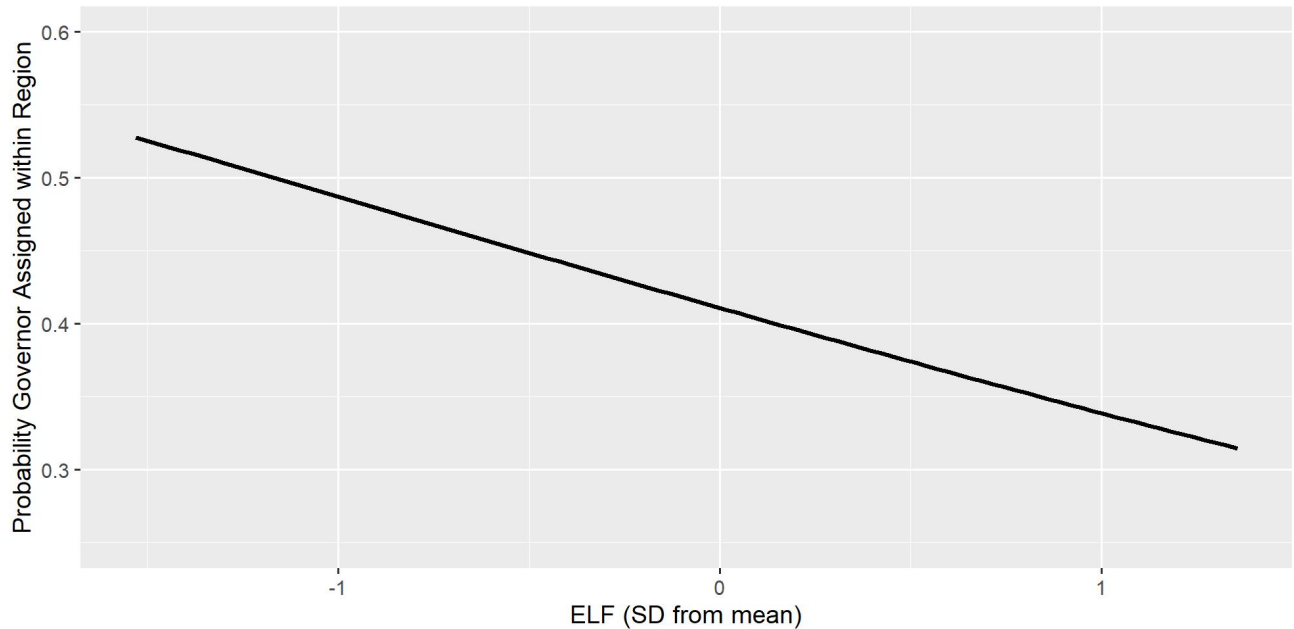
Note: Predicted probabilities calculated based on Model 2 of Table 5.6, and for a province with median Population, Elevation, with Sea Opening, without a Railroad, with Land Border, and in the Anatolia region.

of ethnic Albanian origin and was the de facto ruler (Sezer Feyzioğlu 2017) of large Greek and Albanian-populated territories in the Ottoman Balkans, used Greek in his court.

The difficulties posed by linguistic diversity gave the Ottoman administration several choices. One was to teach its bureaucrats the local languages (Karpat 1985, p.32, fn. 38). However, this is only a strategy that the administration could rely on in the long term. This requires significant investment in opening schools and educating students at these schools. Another long-term strategy was to open schools in the provinces and educate bureaucrats among the populations who already spoke the local language, or to send students from other regions there to be instructed in the local language. The state opened schools in the provinces, providing education in administration and law, to invest in future administrators and judges with a command of local language (i.e., BOA.BEO. 4185/313863). Nevertheless, this is still not ideal to achieve a high-capacity state under a diverse population. These bureaucrats could only be assigned to regions where the lan-

guage they learn is spoken, but not anywhere else. This and the fact many of them were already from these local populations risk the possibility that they will build local patronage relationships undermine the state's control and capacity.

Another strategy was to narrow the pool of candidates for the job and send those who could speak the local language. Yet, I argue that such a strategy also jeopardizes the state's capacity as it prevented the optimal bureaucratic appointments, constraining who could be appointed where.

A civil service examination is one strategy to hire more skilled bureaucrats to increase the state's capacity. From a telegram sent from the center to the Yemen governorate, we learn that the Ottoman administrators in Yemen had attempted to implement a province-wise examination for civil servants, and those who failed this examination would not be able to keep their jobs (BOA.DH.MKT. 1838/16). Apparently, many exam-takers failed the exam. In this telegram the central government asks the governor of Yemen to be more tolerant *(müsamahakar)* towards those who failed the exam if they could speak the local language (which is Arabic) and that they should not be dismissed from their position without explicit approval from the center. Here, we have a clear case where a minority language constrains the state's bureaucratic capacity in a minority-populated province. In Yemen, the vast majority of the population spoke Arabic. The bureaucrats who were not competent enough for their jobs, those who failed the examination, could keep their job only because they spoke Arabic. The lack of Arabic-speaking civil servants in the empire forces the government to keep civil servants who are not sufficiently competent in their position.

The Balkan provinces of the empire were linguistically and religiously very diverse. This diversity undermined bureaucratic capacity in these regions by limiting the number of possible candidates for the open positions. One example I will discuss here is again from the year 1907. This is a letter written to the Office of the Prime Minister *(Sadaret)* by the Rumeli Inspectorship *(Rumeli Müfettişliği)*, where the bureaucrats from the inspectorship inform the Prime Minister that in the provinces of Monastir *(Manastır)*, Kosovo *(Kosova)* and Thessaloniki *(Selanik)* the new district governors to be appointed under the inspectorship's area of jurisdiction would be required to speak

French in addition to having good command of one local language (BOA.TFR.I.A 36/3546).[2] These are high standards for a district governor. What further aggravates the issue is that it would still be not ideal if the district governor spoke only one of the local languages. In these three provinces there were large communities who spoke Albanian, Bulgarian, Greek and Serbian, in addition to Turkish. A district governor who only speaks one of these local languages (in addition to French) can only efficiently serve in another district where the same local language is spoken. Assigning this governor to a district where the majority of the population speaks another language could undermine bureaucratic capacity due to linguistic barriers. Not being able to assign any civil servant anywhere where they can serve the state better does not yield ideals conditions where the state can improve administrative as well as fiscal capacity because if an administrator serves in an area for long periods of time it invites the possibility of patronage-ridden relationships between the locals and the administrator, to the detriment of the central state.

A similar phenomenon in Yemen in the year 1888 presents another case where the state's bureaucratic capacity suffered due to lack of bureaucrats who spoke the minority languages. In this telegram, it is noted that the kaza governor *(kaimmakam)* of the Radaa *(Rada)* kaza (third-level administrative district) has left for Istanbul and resigned his post. The main purpose of this telegram is to inform the recipient that a correspondence clerk of second-class *(ikinci sınıf tahrirat katibi)* named Mustafa Fehmi Efendi with sufficient knowledge of the Arabic language and customs *(lisan ve mizac-ı Arab'a vakıf)* has been assigned as the kaza governor here since no assistant kaza governors from other regions with a command of the Arabic language and customs accepted this post. This is likely to deteriorate bureaucratic capacity. Because in the Yemen province an overwhelming majority of the population spoke Arabic, the center wanted to assign a kaza governor who spoke Arabic and was familiar with the Arab customs. Such credentials could make it more likely for the administrator control and more efficiently rule the population compared to someone without these credentials. However, these criteria severely narrow the number of potential candidates for the job.

---

[2]The Rumeli Inspectorship aimed to implement reforms in three Balkan provinces, Monastir, Kosovo and Thessaloniki. The requirement for speaking French was probably not due to diversity, but in order to be able to communicate with foreign diplomats in the region.

When all of the potential candidates who are even by then assistant kaza governors, already lower in the ranks and probably less experienced than kaza governors, refuse the job, the center has to assign a correspondence clerk to the position of kaza governorship. It is likely that someone from such lower ranks of the administration has less experience and less competence for the job, which is not ideal for bureaucratic capacity.

The problem was not only that bureaucrats could not speak the minority languages, but also that some bureaucrats of minority backgrounds could not speak Turkish, the state language, well enough. Muhammed Hilal Efendi, whose report about Yemen I discussed in Chapter 4 was born in Aleppo, and apparently was of Arab origin (BOA.DH.SAİD.d. 4/1032; Göyünç 1980). After having served in Yemen and Tripoli of Libya, which are Arab-speaking provinces, he was assigned to Konya, in today's central Turkey, which was a majority Turkish province. However, according to the state records, he could not efficiently perform his duty in the Court of Appeals *(İstinaf Mahkemesi)* in this province since he had problems in understanding and speaking the Turkish language, being "from the people of Arabia" *(Arabistan ahalisinden olmasıyla)*. Because of these problems, the records indicate that it was decided he would be better off serving in Arab lands and his appointment in Konya was terminated (BOA.DH.SAİD.d. 4/1032; Göyünç 1980).[3]

The necessity of appointing bureaucrats with knowledge of the language and customs of the area is a recurring theme across many sources. In his report, Muhammed Hilal Efendi recommended the Sultan to appoint officials who spoke Arabic and were familiar with Arab customs to Yemen, or even better appoint Arabs (BOA.YEE. 58/33). Secondary sources on Ottoman governor appointment patterns confirm how the Ottoman state had to appoint bureaucrats across provinces that had ethnic demographics which resembled each other. This can be clearly observed with the within-region bureaucratic appointment patterns in the Arab regions of the empire (Anscombe 1997; Çetinsaya 2006).

Çetinsaya (2006) mentions that the governorships of Mosul and Baghdad provinces were usually given to the bureaucrats who had already served in these areas as they knew the area and

---

[3]From the report he wrote, it is clear that Muhammed Hilal Efendi was competent in written Turkish. The problem here probably was only in his speaking and understanding spoken Turkish.

possessed strong networks. Regarding the appointment patterns of district governors *(mutasarrıf)* of the Najd *(Necd)* sanjak in 'Eastern Arabia' during the 1870s, Anscombe mentions several examples of mutasarrıfs assigned to this region. These include Said Bey, an Arabic-speaking army official whose previous service was in Basra; former district governor of Hudayda *(Hudeyde)* in Yemen Ibrahim Pasha; former district governor of Asir in Yemen, Musa Kazım Efendi; Said Pasha, who was from Baghdad and who had served in Hasa in Eastern Arabia for thirteen year; Sayyid Talib Pasha who was from Basra; and finally Sayyid Akif Pasha. In addition to this, the military commander in Iraq, Nafiz Pasha had become the district governor while "one of his leading officers" (1997, p.64) was assigned as the deputy district governor and the treasury accountant had previously served in Kirkuk *(Şehr-i Zor)*, in Northern Iraq and his treasury accountant was previously the accountant of the same place. In the same period of time the district governor of Qatif in Eastern Saudi Arabia, had previously been a district governor of Samarra in Iraq, and the district governor of Mubarraz, again in Eastern Saudi Arabia had formerly been a treasury official in Qatif (Anscombe 1997, p.50). Anscombe explains that the reason for these assignments was that "[T]he state wanted representatives who knew local customs, were not cut off from the populace, and had the experience to counter domestic and foreign security threats" (pp.64-5).

5.3 Conclusion

Where the populations are diverse, the different languages and customs of the population constrain the state's bureaucratic capacity because they impose constraints on the state's ability to make the optimal bureaucratic assignments. In this chapter, using a dataset on the governor assignments in late Ottoman Empire and also using archival sources, I demonstrated that ethnolinguistic diversity affects bureaucratic appointment patterns.

The large-N dataset I used for the empirical analysis in this chapter was only for bureaucratic appointments of governors for the provinces, the highest-level administrative unit in the empire. Even though I utilized evidence from archival documents to demonstrate that similar patterns existed at the lower-levels of the bureaucracy, for future data collection efforts it would make more

sense to collect large-N data on the bureaucratic appointments at the lower-levels of the Ottoman provincial administration.

More importantly, while with the evidence I use here I can demonstrate that diversity restricts bureaucratic appointment patterns, I cannot provide any direct evidence on bureaucratic capacity as I lack any data where I can measure local-level bureaucratic capacity in the Ottoman Empire during this period. This can be the second avenue for future work, to construct a dataset where local-level bureaucratic capacity can be measured.

Chapter 6: Taxing Diverse Populations: Evidence from Local-Level Taxation Patterns

On June 4th 2021, in a context when corruption among high-profile Turkish politicians and bureaucrats was hotly debated among the Turkish public, Rüşdü Saracoğlu, a former Central Bank Governor and a former minister of Turkey sent a tweet where he praised his grandfather Şükrü Saracoğlu, who had served as a prime minister of Turkey for being an honest politician, noting that he had died without being a homeowner.

This tweet generated significant controversy because his grandfather Şükrü Saracoğlu was the prime minister during the latter part of World War II, between 1942 and 1946, and he is seen responsible for implementing horrendous anti-minority policies. One of these is the "Varlık Vergisi" (Wealth Tax) of 1942 which almost exclusively targeted non-Muslim minorities and was devastating for them. Many non-Muslim citizens were levied exorbitant amount of taxes that often were far higher than their total wealth. Many of them lost their homes and businesses, and those who could not pay the tax were sent to labor camps to build railroads in Eastern Turkey where many died (Aktar 2001; Bali 2005).

Many replies to Saracoğlu's tweet reminded him that his grandfather, whom he praised, was the prime minister during the implementation of this tax and is often seen as its chief architect (Ökte 1987). To one of these tweets he responded by defending the tax, arguing that who had enjoyed privileged status during the Ottoman era had to pay their fair share of taxes in the Turkish Republic (Saracoğlu 2021).[1]

This reflects a very common perception among the Turkish elite and the public, that the minorities were not paying their fair share in taxes and military service in the Ottoman Empire, which

---

[1]He tweeted in Turkish: "...Senelerce Osmanlının kaymağını yiyenler bir gün Cumhuriyette bunun vergisini vermeliler değil mi?"

had enabled them to become wealthier than the Turkish Muslim populations. The violent episodes of ethnic cleansing of the minority populations in late Ottoman Empire and early Turkish Republic were at least partly based on these perceptions and aimed to transfer wealth from the non-Muslim populations to the Muslim populations (Aktar 2001).

In this chapter, I examine the conditions that created the unequal burdens across the dominant group, Sunni Muslim Turks, and the minorities in late Ottoman Empire. I empirically test my expectations on how diversity ended up impeding state building by discouraging investment in state capacity and how it resulted in the tax burdens being placed on the dominant group in the society, Sunni Muslim Turkish speakers.

The chapter is organized as follows. Section 6.1 describes the data and the method I use in the quantitative empirical analysis using the local-level tax revenue data, and presents the results of these analyses. Section 6.2 presents evidence from individual-level property surveys conducted in the 1840s. Finally, Section 6.3 presents and discusses additional qualitative evidence for the mechanisms I offer from correspondences among the Ottoman administrators in the Ottoman archives.

6.1 Evidence from Local-Level Tax Revenue Data

6.1.1 Data and Variables

Commonly used measures for fiscal capacity concern how much tax revenues the state can extract (Levi 1988; Cheibub 1998). I use tax revenues per capita as a measure of fiscal capacity. In order to construct this measure I collected province- sancak-level revenue data from archival and historical sources and estimates of the province-level population.[2] Since revenues can change according to prices, adjusting for prices is also necessary. I adjust the revenues according to the index by Pamuk (n.d.). Using his index, I convert the revenues to their equivalents in 1998 US Dollars. Therefore, the dependent variable in the empirical analysis is Revenue per Capita in 1998 US Dollars.

---

[2]Province *(vilayet)* is the first-level administrative unit in the empire during this period, and sancak is the second-level administrative unit.

I constructed the historical local-level tax revenue data from hundreds of historical and archival sources mostly in the Ottoman Turkish language in order to measure fiscal capacity. The main part of the archival research in the Ottoman Archives in Istanbul took more than 18 months. The sources I use include archival sources which exist under several classifications in the Ottoman archives and the Provincial Yearbooks *(Vilayet Salnameleri)* published by the Ottoman provincial administrations. The latter include budget reports sent to the center by provincial administrators and budget reports prepared by bureaucrats in the center. In addition to these, I relied on several other sources such as the British Foreign Office reports, statistical yearbooks published by the Ottoman central state and secondary sources by historians.

The first known provincial yearbook in the Ottoman Empire is published for the Ottoman fiscal year (Rumi year) of 1283 (1867/1868 in Gregorian calendar) by Bosnia Province (McCarthy and Hyde 1979). The yearbooks were not published very consistently by all provinces. The frequency of the publication of the provincial yearbooks seems to decrease in early twentieth century, and they disappear with some exceptions in the 1910s, leaving their place to yearbooks and other statistical sources published by the central state.[3]

The first major source of revenue data is the provincial yearbooks. I have been able to identify around 550 provincial yearbooks and have been able to find around 500 of them. An average yearbook consists of a few hundred pages. Most of the yearbooks contain information on the total fiscal revenues at the province level. However, in some of them the information is incomplete (for example, they report revenues from only certain types of taxes) and some report duplicate information with other years. Some yearbooks also contain very detailed information about different type of taxes and report tax revenues at lower-level administrative districts, including sancaks.

A second major source of revenue data is the archival documents I could identify in the Ottoman Archives in Istanbul. These documents include budget reports prepared by the administrators in the provinces and sent to the center and others are budget reports prepared in the center. These documents were scattered among different fonds in the Ottoman archives. The picture of one of

---

[3]McCarthy and Hyde (1979) provide a list of published provincial yearbooks they could identify.

77

Figure 6.1: Example Archival Document Showing Local-Level Fiscal Revenues, Kosovo Province for the Fiscal Year 1320 (1904/1905).

Source: BOA.TFR.KV.84/8377

these documents is in Figure 6.1. It is a report that provides the detailed revenue and spending items in the Kosovo province, for the fiscal year 1320 (1904/1905 in the Gregorian calendar), for second-level (sancak) and third-level (kaza) administrative units. Kosovo province includes territories in today's North Macedonia, Serbia, Albania, Kosovo and Montenegro.

In addition to these two major sources, I also relied on several others that provide a compilation of province-level tax revenues for some years. One of them is the statistical yearbook published by the Ottoman state in the year 1897 (Güran 1997). It provides information on tax revenues in all Ottoman provinces for the fiscal year 1310 (in Gregorian calendar 1894/1895). The second such compilation is the Fiscal Statistics Journal *(İhsaiyat-ı Maliye)* published by the Ottoman Ministry of Finance in fiscal year 1327 (Gregorian years 1911/1912) which provides province-level tax revenues for the fiscal years 1324 and 1325 (Gregorian years 1908/1909 and 1909/1910). Another

compiled source published by the Ottoman central state is the General Fiscal Statistics *(Hazine Hesab-ı Umumisi)* in the Hicri year 1330 (Gregorian years 1911/1912) which provides province-level tax revenues for the fiscal year 1326 (Gregorian years 1910/1911). For the pre-1877 observations, the British Foreign Office documents (Harrison 1984) provide a compilation of the revenues of all provinces. Finally, I used data from numerous secondary sources such as dissertations and published articles.

Since I used many different sources, for some units in some years there were two or more sources which provided fiscal revenue data. For the cases this happened, I constructed the following hierarchical criteria to decide which source to use. If the data is a compilation of all the provinces in a published source by the Ottoman state I used this. Next come compilations of all the provinces in other state documents that report more than three provinces (i.e. reports in the Ottoman archives) and foreign sources (i.e. the British Foreign Office documents). The third factor I consider in the hierarchy is whether the data reports revenue items. I prioritized data that reports the items over those that only report the total revenue. Fourthly, I choose the data in the Provincial yearbooks. Fifth, I use the documents I found in the Ottoman Archives that report the data for at most three provinces. Finally, at the bottom of the hierarchy is the data reported in secondary sources that were published.

Even though some of these sources contain tax revenue data at the lower-levels of administrative units, for the main empirical analyses I use the province-level data since lower-level data are much scarce and do not adequately cover all the periods I am interested in to allow for statistical analyses. Because the sancak-level data is more limited, I conduct the analyses with the sancak-level data, only as an extra layer of analyses to check whether similar patterns can be found at this level.

To calculate per capita tax revenues at the province level, I needed the total population estimates for each province. I reconstruct local-level population data from multiple sources. The main sources I used for the total population in a province are Ottoman censuses. Unless a province's censuses were unavailable, or unreliable, I used the population data from the Ottoman censuses

Table 6.1: Descriptive Statistics at the Province-Year Level

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Tax Revenue	750	51.69	25.66	0.12	32.47	49.71	71.71	151.74
ELF	750	0.38	0.36	0.00	0.01	0.31	0.76	0.97
Percent Turkish	750	0.39	0.26	0.00	0.11	0.38	0.66	0.75

to measure the total population in a province. In order to take into account the territories that the Ottoman Empire lost during this period in the Balkans and also to take into account the population changes caused by frequent shuffling lower-level administrative units across provinces in the Balkans, I relied on Akarlı (1972). For the total populations of the Edirne, Thessaloniki *(Selanik)*, Monastir *(Manastır)*, Ioannina *(Yanya)*, Kosovo *(Kosova)* and Mediterranean Islands *(Cezayir-i Bahr-i Sefid)* provinces I used his estimates. For the other provinces, when they had no census or it was unreliable, I used the estimates provided by Karpat (1985) and the British census data as reported by Zamir (1981).

In order to calculate annual population changes in these provinces, I calculated the average of the percentage of annual changes in the empire's total population according to the annual estimates in Shaw (1978). I assumed that the average annual percentage change in the empire's total population was equal to the annual percentage change in each province each year. To deal with jurisdictional boundary changes that would have changed the total population of the province, I referred to the list of all jurisdictional boundary changes (except for those in the Balkans, for which I use Akarlı (1972) as I mentioned above) in Sezen (2017) and adjusted the population estimates of each province according to the changes in jurisdictional boundaries.

As the independent variables, I again use the two measures of diversity I described in Chapter 3, ELF for heterogeneity and Proportion of Turkish population for similarity. Table 6.1 provides the descriptive statistics of the dependent and independent variables at the province-year level and Table 6.2 provides the descriptive statistics at the province level (averaged for each province across all years).

Table 6.2: Descriptive Statistics at the Province Level

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Tax Revenue	29	48.51	24.14	0.54	31.43	40.97	69.01	87.80
ELF	29	0.33	0.35	0.00	0.01	0.16	0.75	0.97
Percent Turkish	29	0.39	0.27	0.00	0.11	0.39	0.63	0.75

The population data for the sancaks are more limited. The population statistics for ethnolinguistic identity cannot be reconstructed for many sancaks. In order to calculate the ELF and Percent Turkish measures at the sancak level, I rely mostly on the Ottoman censuses, and the British Population Statistics (Zamir 1981).

6.1.2 Identification Strategy

I argue that for a study of fiscal capacity building a within-country design is preferable over a cross-national design because wihin-country studies can help address several problems that cross-national studies cannot.[4] Most importantly, by considering the tax revenues of a single country, I can hold constant an unobservable factor that renders cross-national measurement of fiscal capacity very problematic. The most commonly used measure of fiscal capacity, tax revenue per capita, can be problematic in cross-national designs since the amount of taxes that the state can extract can also be affected by factors other than the state's capacity to tax, such as the willingness of the government to tax. A country can have low tax revenues even though it has the capacity to tax, because the government just does not (want to or need to) tax. Alternatively, it can have low tax revenues even though the government wants to increase tax revenues because it lacks the capacity to tax. It is more difficult to distinguish between these two alternative scenarios in cross-national studies of fiscal capacity. A single-country design helps address this problem by holding constant the ruler's need for funds.

There are further advantages of restricting the analysis to a single country. It can help control for any other unobservable factors that can confound the relationship between the variables of

---

[4]See Lieberman (2002) for a discussion of cross-national designs of studies with taxation data.

interest, can eliminate the commonly criticized measurement errors in cross-national studies and also alleviate the concerns regarding the different effects—even the same— war can create in different cases. Because war is the critical factor here it is important to restrict the scope of the analysis to a single country. It could have been problematic to examine war's effect on different countries since even the same war can have different effects on different countries. Examining war's effect on different regions of a single country, and therefore on a single ruler with a single treasury, can help us estimate this effect without worrying about country-level confounders.

I leverage the shocks created by the two interstate wars that the Ottoman Empire fought in this period. The first is the 1877-8 Russo-Ottoman war, which lasted almost a year, was fought on multiple fronts, was devastating for the empire, and resulted in utter defeat and significant loss of territories and populations. The second war happened two decades later. It is the 1897 Greco-Ottoman war, which was much shorter (just over a month), fought on a much smaller geographic area and resulted in Ottoman military victory. This can also allow us to see if the scales and lengths of these two wars made any difference. In the empirical analysis I describe below, I estimate how the diversity measures change the differences in the fiscal revenues before and after each war.

I use a *generalized difference-in-differences* model in order to evaluate whether changes in fiscal revenues across the wars are lower in less diverse administrative units as I hypothesized. Both *ELF* and *ES* variables are continuous. This implies that I have a generalized difference-in-differences model with continuous so-called 'treatments' because their interactions with the war dummies are also continuous. I estimate an Ordinary Least Squares regression model in the following form:

$$Revenue_{it} = \beta_0 + \beta_1 ELF_{it} + \beta_2 ES_{it} + \beta_3 D_t^{1877} + \beta_4 D_t^{1897} + \beta_5 D_t^{1877} ELF_{it}$$

$$+ \beta_6 D_t^{1897} ELF_{it} + \beta_7 D_t^{1877} ES_{it} + \beta_8 D_t^{1897} ES_{it} + \lambda_t + \gamma_i + \varepsilon_{it}$$

In this specification $Revenue_{it}$ is tax revenue per capita (in 1998 USD) in province i in year t. ELF_{it} is the Etholinguistic Fractionalization measure, and ES_{it} is the ethnic similarity measure (Percent Turkish population) in administrative unit i and year t. D_t^{1877} and D_t^{1897} are dummies for

82

post-1877 and post-1897 periods respectively. $D_t^{1877} ELF_{it}$ and $D_t^{1877} ES_{it}$ are the interactions between the post-1877 dummy and the ethnic composition measures, and $D_t^{1897} ELF_{it}$ and $D_t^{1897} ES_{it}$ are the interactions between the post-1897 dummy and the ethnic composition measures. Finally, λ_t denotes the unit fixed effects, γ_i denotes the province fixed effects and ε_{it} is the error term. The unit fixed effects help account for any year or unit-specific and time-invariant source of heterogeneity. The year fixed effects control for shocks that affect the Ottoman Empire as a whole in certain years.

In this specification, I am interested in the coefficients of the four interaction terms, β_5, β_6, β_7 and β_8 as they measure the difference-in-differences estimates according to the ethnic compositions of the administrative units. I expect β_5 and β_6 to be negative (since I expect heterogeneous provinces to have lower increases in capacity) and I expect β_7 and β_8 to be positive (since I expect similar provinces to have higher increases in capacity).

I had to exclude some provinces from this analysis. Some of these provinces exist for a very short time within the time frame that this study is interested in and therefore they have very few observations. These are Bosnia *(Bosna)*, where Ottomans lost control after the 1878 Treaty of Berlin, and Prizren, a short-lived province for which there is only one observation in the dataset. I also excluded the province of Crete (*Girit*), which left Ottoman control in 1898, because this province had a somehow autonomous status during the period of this analysis. Finally, I do not include the independent second-level administrative units *(bağımsız sancaklar)*.[5] Figure 6.2 below is a presentation for which year in which province the revenue data is available or missing at the province-level dataset. It also shows whether a province existed in a given year.

6.1.3 Parallel Trends Assumption

One worry about the validity of the results in such a difference-in-differences study on wartime increases in fiscal capacity can be the violation of the parallel trends assumption. A certain type of non-parallel trends in this case could be driving the results. This would look like a case where, for

---

[5]Normally, second-level administrative units, sancaks, were under the jurisdiction of a higher-level administrative unit, province (vilayet). However, several sancaks had special status and were not under the jurisdiction of a province.

Figure 6.2: Tax Revenue Data by Province and Year
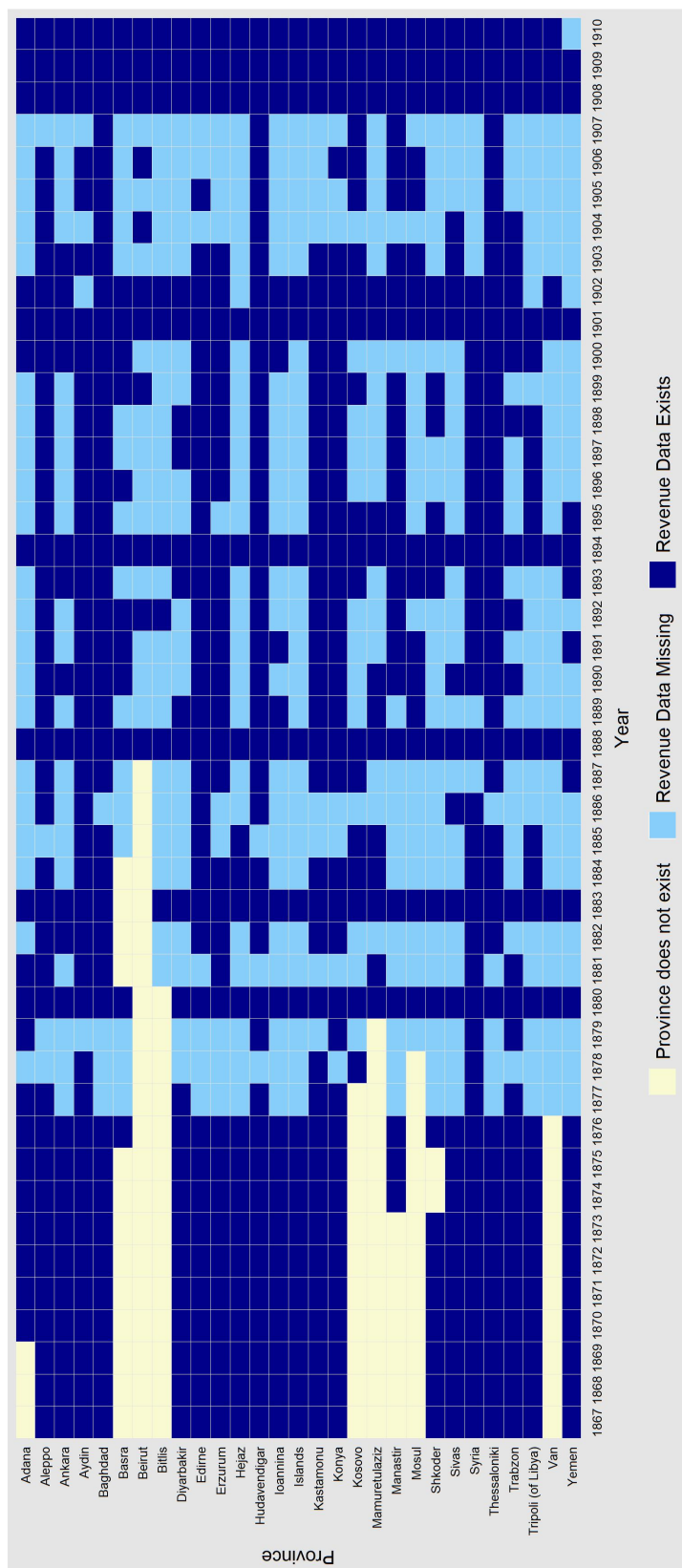
Note: Produced using panelView package in R (Mou, Liu, and Xu 2022)
.

Table 6.3: Pre-War Trends in the Revenues by ELF and ES Variables

	Dependent Variable:	
	Revenue Per Capita (in 1998 USD)	
	Pre-1877	1879-1897
	(1)	(2)
Ethnolinguistic Fractionalization (ELF)	24.957***	19.606***
	(1.760)	(2.672)
Percent Turkish	10.753***	9.151***
	(1.821)	(2.490)
Year Count	−0.152	−0.056
	(0.188)	(0.088)
ELF * Year Count	−0.239	0.003
	(0.188)	(0.088)
Percent Turkish * Year Count	0.097	0.050
	(0.178)	(0.088)
Province Fixed Effects	Yes	Yes
Observations	218	297
R^2	0.929	0.899
Adjusted R^2	0.920	0.887

Note: OLS Regression. Standard errors in parentheses. $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

example, the per capita revenues in more homogeneous and more similar provinces were already rising at a higher pace compared to the less homogeneous and less similar ones before the two wars happened and kept rising at a higher pace after the war so that the wartime increases end up to be higher in the more homogeneous and similar provinces.

Even though it is not formally entirely possible to test the parallel trends assumption, I conducted an analysis which aims to show that the annual changes in the dependent variable before each of the wars are not linearly correlated with any of the two explanatory variables. I simply interacted a year count variable with each of the explanatory variables. If the coefficients of these interaction terms are estimated to be statistically significant, this can be an indication of a violation of the parallel trends assumption.

I present the results in Table 6.3. Model 1 tests the trends for the years leading up to the war of 1877-78 from the starting point of the dataset, 1867 and Model 2 tests the trends for the years after

the 1877-78 war and before the 1897 war. The results indicate that there is no pre-war trend that is correlated with either of the explanatory variables that can cause any concern. The coefficients of the interactions of the explanatory variables with the Year Count variable are neither substantively, nor statistically significant.

6.1.4   Results at the Province Level

Table 6.4 reports the results of the generalized difference-in-differences regression and Figure 6.3 presents the same results visually, in the form of a coefficient plot. The coefficients of interest are the interaction effects between the period dummies and the diversity measures. Both ELF and Percent Turkish variables are standardized around the mean so that the coefficients of the main effects of ELF and ES refer to the estimated change in the fiscal revenues per one standard deviation increase in these variables, and the coefficients of the interaction effects refer to the changes in the changes in the fiscal revenues after each war per one standard deviation increase in the independent variable.

The size of the coefficient of *ELF * Post-1877 Dummy* interaction is -2.010 and is significantly estimated. This suggests that per one standard deviation (0.26) increase in the province's ELF measure, the estimated decrease in the change that 1877-78 war causes is 2.01 US Dollars per per capita. Next, checking the coefficient of the *ELF * Post-1897 Dummy*, one can see that it is estimated to be negative at -2.087. This suggests that one standard deviation increase in ELF decreases the change caused by the 1897 war by 2.087 US Dollars per capita, and this coefficient is estimated to be statistically significant. These results, are in line with my expectations and indicate that the increase in the Ottoman state's fiscal revenues were lower in more heterogeneous provinces across each war.

To put these numbers in context, we can compare the changes across different values of the ELF measure before and after wars happen. For example, the estimated coefficient for the Post-1877 Dummy in model 1 is 5.556. This is the estimated increase that happens after 1877 for a country that has the mean ELF value (0.39). The coefficient of the interaction between this

Table 6.4: Generalized Difference-in-Differences Analysis of the Wartime Increases in Fiscal Revenues with Province-Level Data

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	22.338***
	(1.209)
Percent Turkish	11.029***
	(1.162)
Post-1877 Dummy	5.556*
	(3.124)
Post-1897 Dummy	6.266**
	(2.910)
ELF * Post-1877 Dummy	−2.010**
	(0.812)
ELF * Post-1897 Dummy	−2.087***
	(0.773)
Percent Turkish * Post-1877 Dummy	3.129***
	(0.805)
Percent Turkish * Post-1897 Dummy	3.537***
	(0.784)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	747
R^2	0.904
Adjusted R^2	0.893

Note: OLS Regression. Standard errors in parentheses.　　　*p<0.1; **p<0.05; ***p<0.01

dummy and ELF being -2.01 suggests that the estimated change after 1877 in a province one standard deviation below mean ELF (which has an ELF of 0.13) is the difference between 5.556 and -2.01 and therefore 7.566 USD per capita. On the other hand, in a province with an ELF measure one standard deviation above the mean (with an ELF of 0.65), this change is estimated to be 3.546. Therefore, the change that happens after 1877 is estimated to be almost 2.1 times as large in a province that has an ELF value one standard deviation above the mean compared to a province with an ELF one standard deviation below the mean.

Next, I check the coefficients of the *Percent Turkish * Post-1877 Dummy* and *Percent Turkish * Post-1897 Dummy* interactions to evaluate whether the expectations also hold for the *Ethnic Simi-*

Figure 6.3: Coefficient Plot for the Difference-in-Differences Analysis
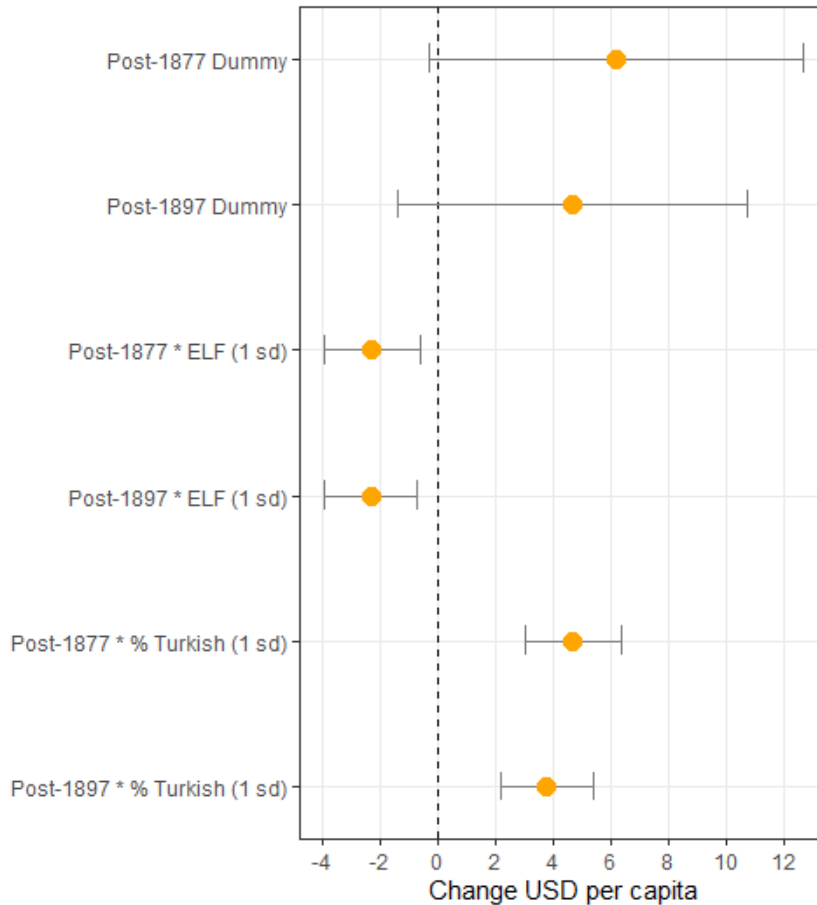
larity (Percent Turkish) variable. The coefficient for the Percent Turkish * Post-1877 Dummy interaction is 3.129 and is significantly estimated. This suggests that there is a 3.129 per capita USD increase in the changes in fiscal revenues per one standard deviation increase (0.36) in the percentage of Turks, after 1877. Finally, the coefficient for the Percent Turkish * Post-1897 Dummy interaction is 3.765 and this coefficient also is estimated to be significantly different from 0. Hence, the change in the changes in fiscal revenues after 1897 is 3.537 per capita USD higher per one standard deviation increase (0.36) in the percentage of Turks. These results, are again consistent with the expectations. The wartime changes in fiscal revenues are higher in more ethnically similar administrative units, which had higher proportion of Muslim Turkish speakers.

An important note to make here is is the relationship between diversity and the absolute levels of per capita fiscal revenues. I do not discuss these relationships in evaluating the hypotheses. Even

though the argument I put forward in this study implies that *every other factor held constant*, in more diverse places taxpayers should be paying lower taxes, the main hypotheses I presented are about wartime changes in fiscal revenues. I do not focus much on the overall levels of tax revenues per capita because it would be misleading to compare the revenues across different provinces in this case, having no reliable GDP estimate for each province for each year and therefore not being able to take GDP into account.

The findings in Table 6.4 demonstrate that overall per capita revenues are higher in provinces with higher values of the ELF, indicating taxpayers in more heterogeneous provinces on average pay more taxes. Per capita revenues are also higher in provinces with higher values of ES, indicating Turks on avarage pay higher taxes per capita than other ethnolinguistic groups. These patterns regarding the ELF, I argue, are likely because I cannot account for the levels of GDP. The Balkan provinces of the Empire have very high ELF scores, and also very high tax revenues compared to many other regions in the Empire. There were many important port cities in the Balkans, boosting exports and commercial activities. Income from port cities and the region in their surroundings were important contributors to the Ottoman treasury (Kasaba 1994). The proximity of the Balkan provinces to Europe also contributed to their export potential and volume of trade which should have increased the tax revenues here. The export potential and its effect on trade likely render heterogeneous Balkans have higher per capita tax revenues compared to Arab and Central Anatolian provinces, which are very homogeneous and have lower per capita tax revenues.[6]

There is also no evidence that the tax rates were higher in the Balkans than in the other areas. Discussing the Property Censuses *(Temettüat sayımları)* of the 1840s, Güran (2014, p.316) writes that the tax rates were lower in the Balkans than in other areas of the empire. In any case, I will try to address this issue, by analyzing individual-level taxation data from Property Registers *(Temettüat Defterleri)* from the 1840s. Doing this, I can focus on a certain settlement, thereby keeping constant the potential for economic output, and can check whether the individual members of the ruling groups on average paid more taxes compared to minorities.

---

[6]Table 5.1 in Chapter 5 reported the average within-region ELF scores.

6.1.5 Results at the Sancak Level

The Ottoman provinces were the highest-level administrative units. It is possible that there was variation in the population's diversity and the state's capacity within provinces and the analyses I presented above may be suffering from the problem of ecological inference.

In order to alleviate these concerns, I replicate the analyses at the sancak level. This analysis includes the independent sancaks, the second-level administrative units that are not under the jurisdiction of any province. These independent sancaks are Benghazi *(Bingazi)*, Biga/Kale-i Sultaniye, Çatalca, İzmit, Jerusalem *(Kudüs)*, and Zor.[7]

This analysis is imperfect in terms of missing data due to three reasons. The first reason why these data have problems is, as I discussed in Chapter 4, data on the population of ethnolinguistic groups do not exist for many sancaks. The second reason why there is a problem of missingness in this data is that for many sancaks there is no tax revenue data at all, or there is no tax revenue data for multiple periods. Hence, I lose many of the observations and can conduct the analysis only in a small subset of the sancaks. The third and final reason is the jurisdictional boundary changes, where some kazas were assigned from the jurisdiction of one sancak to another at certain points, which results in changes in their total level of due to the changes in their total population. Based on the jurisdictional boundary changes took place according to Sezen (2017), I adjusted the population estimates for some sancaks; however, for some other sancaks I could not reconstruct population estimates because some kaza-level population estimates are not available. This resulted in some sancaks completely being dropped from the sample, while for some others data is dropped for some periods. I provide a list of the sancaks and the time periods for which they have data in Appendix Table C.1.

The dependent variable is inflation-adjusted total revenues at the sancak level in million guruşes. I adjust the revenues according to the price index of Pamuk (n.d.), and the revenues are million USDs in 1998 value. Because reliable total population estimates are not available for all

---

[7]These units were not always indepdendent sancaks, at certain periods they were regular sancaks under a jurisdiction of a province.

of the sancaks, I do not calculate the per capita revenues for this analysis as it could result in losing further observations. Due to the lack of sufficient within-unit observations across different years, I take within period (Pre-1877, 1877-1896, and Post-1897) averages of revenues for each sancak. I standardize the explanatory variables around their mean, so that their coefficient refers to the estimated change in the dependent variable, revenues, per one standard deviation increase in the explanatory variable. Similar to the province-level analysis, I estimate the following OLS Regression:

$$Revenue_{it} = \beta_0 + \beta_1 ELF_{it} + \beta_2 ES_{it} + \beta_3 D_t^{1877} + \beta_4 D_t^{1897} + \beta_5 D_t^{1877} ELF_{it}$$

$$+\beta_6 D_t^{1897} ELF_{it} + \beta_7 D_t^{1877} ES_{it} + \beta_8 D_t^{1897} ES_{it} + \gamma_i + \varepsilon_{it}$$

Note that the difference of this equation from its counterpart that estimates the results for the province-level data is that I cannot include year fixed-effects (λ_t) in this case because each observation has a single revenue value for each period.

I report the results with the sancak-level data in Table 6.5. The results are only weakly, if at all, consistent with Hypotheses 1a and 1b. Even though all the coefficients of the difference-in-differences estimates (interaction terms between the diversity variables and the period dummies) are in the expected direction, none of them is estimated to be statistically significant. Checking the sizes of the difference-in-differences coefficients, however, we can see that except for one coefficient, the interaction effect between Percent Turkish and Post-1877 dummy, all are substantively meaningful.

The analysis in Table 6.5 may be problematic because the panel data here is unbalanced, many units lacking observations in some time periods even after when the data is collapsed within each period. I conduct empirical analyses with two different subsets of the data that that includes only two adjacent periods (i.e., one subset that includes the 1868-1876 and 1877-1896 periods, and another subset that includes 1877-1896 and 1897-1910 periods) to examine the changes in fiscal revenues by diversity.

Table 6.5: Difference-in-Differences Analysis of the Wartime Increases in Fiscal Revenues with Sancak-Level Data

	Dependent Variable:
	Total Revenues (Million in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	6.701***
	(2.440)
Percent Turkish	−2.471
	(1.632)
Post-1877 Dummy	0.835***
	(0.268)
Post-1897 Dummy	1.450***
	(0.226)
ELF * Post-1877 Dummy	−0.457
	(0.275)
ELF * Post-1897 Dummy	−0.169
	(0.216)
Percent Turkish * Post-1877	0.018
	(0.296)
Percent Turkish * Post-1897	0.332
	(0.231)
Province Fixed Effects	Yes
Observations	134
R^2	0.971
Adjusted R^2	0.947

*Note: OLS regression. Standard errors in parentheses. Revenues in earlier periods population-adjusted with respect to the jurisdictional boundary changes. *p<0.1; **p<0.05; ***p<0.01.*

Because there are only two periods in each subset, one before and one after the war, the dependent variable I use in these analyses is the percentage point changes in fiscal revenues across the two periods. I estimate a much less demanding OLS Regression model of the form:

$$\Delta Revenue_i = \beta_0 + \beta_1 ELF_i + \beta_2 ES_i + \varepsilon_{it}$$

where $\Delta Revenue_i$ is the percentage point change in fiscal revenues in province i, ELF_i is the median ELF score for province i during these two periods, and ES_i is the median Percent Turkish score for province i during these two periods.

Table 6.6: OLS Regression Analysis of Percentage Point Changes in Fiscal Revenues with Sancak Data, Only Two Periods Included

	Dependent Variable:	
	Percentage Point Change in Fiscal Revenues	
	After 1877-8	After 1897
	(1)	(2)
Ethnolinguistic Fractionalization (ELF)	−0.062*	−0.051
	(0.036)	(0.038)
Percent Turkish	0.037	0.079*
	(0.038)	(0.041)
Constant	0.174***	0.263***
	(0.035)	(0.040)
Sample	1868-1896	1877-1910
Observations	34	45
R^2	0.121	0.124
Adjusted R^2	0.064	0.082

*Note: OLS regression. Standard errors in parentheses. The dependent variable in Model 1 is percentage point change in fiscal revenues after the 1877-1878 war, and the dependent variable in Model 2 is the percentage point change in fiscal revenues after the 1897 war. *p<0.1; **p<0.05; ***p<0.01.*

I report the results in Table 6.6. Model 1 estimates how changes across the 1877-8 war change with respect to the diversity measures. Both coefficients are in the expected direction and substantively meaningful. However, only the coefficient for the ELF measure is estimated to be statistically significant and only at the 0.1 level of significance. To get a better idea about the size of the coefficients, we can compare them to the standard deviation (0.265) of the dependent variable. The coefficient of the ELF variable is estimated to be -0.062, indicating that one standard deviation increase in the ELF variable is associated with 6.2 percentage points decrease in the changes in fiscal revenues across the 1877-8 war. This means that one standard deviation increase in the ELF variable decreases the outcome variable by 0.23 standard deviations. The coefficient of the Percent Turkish variable being 0.037 indicates that one standard deviation increase in this variable increases the outcome variable, changes in fiscal revenues, by 0.14 standard deviations.

The dependent variable in Model 2 Table 6.6 is the percentage point change in the fiscal revenues across the 1897 war, and its standard deviation is 0.258. In this model, the coefficient of the ELF variable is estimated to be -0.051 but is not statistically distinguishable from 0. The coefficient being in the right direction, one standard deviation increase in this variable indicates 0.2 standard deviations increase in the outcome variable. The coefficient of the Percent Turkish variable in this model is estimated to be 0.079, which is in the expected direction, but is estimated to be statistically distinguishable from 0 only at the 0.1 level of significance. One standard deviation increase in this variable is estimated to increase the outcome variable by 0.31 standard deviations, indicating a substantively meaningful change.

These two models provide some, although not very strong, support for Hypotheses 1a and 1b. Again, one note of caution here is that each sample in each model contains observations from only a subset of all the sancaks in the Ottoman Empire. However, I believe that these results are indicative that the patterns where diversity decreases the wartime increases in fiscal revenues also hold at the lower-level administrative units in spite of the low number of observations in each model and the low statistical power. The number of observations in Model 1 is only 34 and it increases to 45 in Model 2.

6.1.6 Analysis with a Combined Diversity Index

In order to replicate the analyses with a single diversity measure that contained information about both similarity and heterogeneity, I constructed an a new diversity index by simply adding the inverse of the ES score (1-ES) to the ELF score. The resulting index theoretically ranges from 0 to 2, higher values indicating higher diversity. I standardize the Diversity Index variable around its mean, so the coefficient of this variable indicates the estimated change in the dependent variable per one standard deviation increase independent variable.

Overall, the results I report in Table 6.7 are consistent with Hypotheses 1a and 1b. The wartime increases in fiscal revenues are lower in provinces that have a higher diversity measure. The results are robust to other indexes that I calculated. They include calculating the new index through the

Table 6.7: Generalized Difference-in-Differences Analysis of the Wartime Increases in Fiscal Revenues with a Combined Diversity Index

	Dependent Variable:
	Revenue Per Male Capita (in 1998 USD)
Diversity Index	−47.458***
	(2.787)
Post-1877 Dummy	6.745**
	(3.184)
Post-1897 Dummy	6.993**
	(2.995)
Diversity Index * Post-1877 Dummy	−3.788***
	(0.838)
Diversity Index * Post-1897 Dummy	−4.853***
	(0.772)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	762
R^2	0.896
Adjusted R^2	0.885

Note: OLS Regression. Standard errors in parentheses.   *p<0.1; **p<0.05; ***p<0.01

Euclidian distance, calculated by $\sqrt{ELF^2 + (1 - ES)^2}$), and through the harmonic mean of the two measures, calculated by $2(ELF^{-1} + (1 - ES)^{-1})^{-1}$. I do not present the results of the analyses with these measures.

6.1.7 Robustness Checks

In this section I present and discuss the results of several robustness checks. I replicated the analyses using different empirical specifications, and using different data or measures. The results are robust to all of the robustness checks I present below.

Ethnic diversity can be shaped by political factors (Bleaney and Dimico 2016; Pardelli and Kustov 2022), including states engaging in resettling populations in response to external threats (McNamee and Zhang 2019). In order to take into a possible endogeneity of diversity to war in this case, such as ethnic composition being affected by and ethnic homogeneity having been achieved via violence during the armed conflicts, I added the previously fought number of wars

that took place in a province as a control variable. In order to construct this variable, I compiled a list of all the interstate conflicts that the Ottoman Empire participated between 1600 and 1868 (the year when the analysis with the revenue data starts) from Clodfelter (2008) and the Correlates of War dataset (Sarkees and Wayman 2010). From the description of these wars in Clodfelter (2008) I determined whether each war was fought within the boundaries of the Ottoman provinces. This variable counts the total number of wars fought within the boundaries of a province. I added this variable as an interactive control variable to the generalized difference-in-differences model. The results I report in Table 6.8 indicate that the findings are robust to the inclusion of this variable.

Religious and sectarian diversity can also undermine legibility (Blaydes 2018). Constructing an Ethnolinguistic Fractionalization measure (ERF), as I described in Chapter 3 measure allows me to take into account important religious and sectarian differences. Some examples for such differences within the same linguistic groups are the Shi'a and Sunni Muslims especially in the province of Baghdad or the presence of Christian Arabs in the Levant, who are all coded as Arabic speakers in the ELF measure.

One important issue regarding this index is the existence of the Alevi groups in Anatolia and Alawite groups in the Levant. They are not Sunni Muslims, so they should be included in a separate category while calculating the ERF measure. However, I could find no census or estimate regarding the populations of these two sizable groups. This is why while I estimated the model with all the provinces in the dataset in Model 1 of Table 6.9 and did not account for the existence of these group, in Model 2 I removed the provinces that have substantial populations of these two groups. These are Adana, Aleppo and Syria provinces for the Alawite population and Ankara, Erzurum, Mamüretülaziz and Sivas provinces for the Alevi population.

Another factor to take into account is that the ERF measure caused high collinearity problems in the model since it is highly correlated with the Percent Turkish measure (Pearson $r = 0.388$ for the whole sample and 0.423 for the sample when the abovementioned provinces are removed). This is why I estimate the models without the Percent Turkish variable and its interactions with the dummies. I am not reporting any results regarding the similarity variable here, because the similar-

Table 6.8: Generalized Difference-in-Differences Analysis of Wartime Changes in Fiscal Revenues, Controlling for Number of Wars that Occurred in Each Province in the Previous Centuries

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	22.311***
	(1.220)
Percent Turkish	11.086***
	(1.167)
Number of Wars in Province	−32.022***
	(2.541)
Post-1877 Dummy	5.787*
	(3.241)
Post-1897 Dummy	6.853**
	(2.988)
ELF * Post-1877 Dummy	−2.030**
	(0.817)
ELF * Post-1897 Dummy	−2.316***
	(0.797)
Percent Turkish * Post-1877 Dummy	3.052***
	(0.817)
Percent Turkish * Post-1897 Dummy	3.364***
	(0.803)
Number of Wars * Post-1877 Dummy	−0.384
	(0.537)
Number of Wars * Post-1897 Dummy	−0.505
	(0.525)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	745
R^2	0.904
Adjusted R^2	0.893

Note: OLS Regression. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

ity measure for ethnoreligious diversity is the same with the similarity measure for ethnolinguistic diversity, Percent Turkish.

The results I present in Table 6.9 demonstrate that the findings are similar when ERF measure is used instead of the ELF measure. The wartime increases in fiscal revenues are also lower in provinces where Ethnoreligious Fractionalization is higher.

Table 6.9: Generalized Difference-in-Differences Analysis of the Wartime Increases in Fiscal Revenues with Ethnoreligious Diversity as the Explanatory Variable

	Dependent Variable:	
	Revenue Per Capita (in 1998 USD)	
	(1)	(2)
Ethnoreligious Fractionalization (ERF)	−73.802***	−115.561***
	(4.257)	(4.669)
Post-1877 Dummy	6.704*	4.217
	(3.522)	(3.978)
Post-1897 Dummy	4.014	−0.491
	(3.283)	(3.740)
ERF * Post-1877 Dummy	−1.993**	−3.429***
	(0.899)	(0.968)
ERF * Post-1897 Dummy	−2.847***	−2.482***
	(0.874)	(0.940)
Province Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	748	556
R^2	0.871	0.895
Adjusted R^2	0.857	0.881

Note: OLS regression. Standard errors in parentheses. Models 1 includes all observations, and Model 2 excludes observations with Alevi populations. *p<0.1; **p<0.05; ***p<0.01.

I also replicate the analyses with a measure of heterogeneity in terms of solely religious communities. In this measure, groups of different ethnolinguistic identities are included under the same group based on their religious identity. The similarity measure I use here is Percent Sunni Muslim, and the heterogeneity measure is Religious Fractionalization (RF).

Because the two explanatory variables in this case are again highly correlated (Pearson $r = -0.673$), I estimate two different models where only one of them is included. The results from Model 1 in Table 6.10 indicate that in provinces with religious fractionalization the wartime increases in fiscal revenues were lower, and the results from Model 2 indicate that in provinces with a higher percentage of Sunni Muslims the wartime increases in fiscal revenues were higher. These

Table 6.10: Generalized Difference-in-Differences Analysis of the Wartime Increases in Fiscal Revenues with Religious Diversity as the Explanatory Variable

	Dependent Variable:			
	Revenue Per Male Capita (in 1998 USD)			
	(1)	(2)	(3)	(4)
Religious Fractionalization (RF)	−43.457***		−72.757***	
	(2.484)		(2.820)	
Percent Sunni Muslim		36.053***		54.799***
		(2.081)		(2.145)
Post-1877 Dummy	6.151*	5.748*	4.491	4.249
	(3.209)	(3.108)	(3.540)	(3.439)
Post-1897 Dummy	6.540**	7.224**	4.712	5.807*
	(2.992)	(2.898)	(3.337)	(3.247)
RF * Post-1877 Dummy	−4.293***		−3.495***	
	(0.830)		(0.865)	
RF * Post-1897 Dummy	−1.631**		−1.688**	
	(0.795)		(0.838)	
Percent Sunni Muslim * Post-1877 Dummy		5.436***		4.447***
		(0.824)		(0.879)
Percent Sunni Muslim * Post-1897 Dummy		2.821***		2.967***
		(0.766)		(0.827)
Province Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	743	743	589	589
R^2	0.898	0.904	0.911	0.916
Adjusted R^2	0.887	0.894	0.900	0.906

*Note: OLS regression. Standard errors in parentheses. Models 1 and 2 include all observations, and Models 3 and 4 exclude observations with Alevi populations. *$p<0.1$; **$p<0.05$; ***$p<0.01$.*

results are still in line with the expectations that diversity, in this case, solely religious diversity, should hinder fiscal capacity building.

Similar to the measure of ERF, I could not take into account the Alawite and Alevi minorities while calculating the RF measure due to lack of population estimates for these groups. This is why I estimate these models excluding the Adana, Aleppo, Syria, Ankara, Erzurum, Mamüretülaziz, and Sivas provinces. The results I report in Models 3 and 4 of Table 6.10 are robust to the exclusion of these provinces.

Table 6.11: Generalized Difference-in-Differences Analysis of Wartime Changes in Fiscal Revenues in Two Separate Models

	Dependent Variable:	
	Revenue Per Capita (in 1998 USD)	
	(1)	(2)
Ethnolinguistic Fractionalization (ELF)	48.157***	
	(2.647)	
Percent Turkish		21.187***
		(1.401)
Post-1877 Dummy	6.729*	6.496*
	(3.517)	(3.367)
Post-1897 Dummy	4.286	4.639
	(3.278)	(3.140)
ELF * Post-1877 Dummy	−2.242**	
	(0.909)	
ELF * Post-1897 Dummy	−2.865***	
	(0.862)	
Percent Turkish * Post-1877 Dummy		4.638***
		(0.868)
Percent Turkish * Post-1897 Dummy		4.057***
		(0.838)
Province Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Observations	748	748
R^2	0.871	0.882
Adjusted R^2	0.857	0.869

Note: OLS Regression. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Share of a certain group can be highly correlated with measures of heterogeneity in some contexts (Kustov and Pardelli 2018). Even though the ELF and Percent Turkish variables are not highly correlated in this dataset (Pearson $r = -0.058$), in order to alleviate any possible concern, I replicated the analysis in two separate models, excluding one of the explanatory variable from each model.[8] In the first model I exclude the *Percent Turkish* variable and in the second model I exclude the *ELF* variable. The results remain robust. I report them in Table 6.11.

---

[8]It is not possible to have a province that is very similar and also very heterogeneous.

It can be important to replicate the analyses using alternative data sources in order to address the possibility of bias in certain data sources driving the results. I recalculated the revenue per capita measures, using the population statistics provided by British officials in 1919, as reported in Zamir (1981). The provinces covered in this source includes all the provinces that the empire had not lost by the onset of World War I (with the exception of Hejaz and Yemen). These are most of the Balkan provinces of the empire, Thessaloniki *(Selanik)*, Shkoder *(İşkodra)*, Kosovo *(Kosova)*, Monastir *(Manastır)*, and Ioannina *(Yanya)*. For these provinces, I rely on the same estimates I used in the original analysis. The results of this analysis, reported in Appendix Table D.1 are not different than the results of the original analysis. There are only small changes in the coefficients.

I also calculated the revenue per capita measure using alternative revenue data and replicated the analyses with these data. For the pre-1877 observations, where I use the revenue data from the British Foreign Office sources, I substituted each observation where I had data from an alternative source (either from a provincial yearbook or a report in the Ottoman archives, whichever was available, and using the hierarchy I described above). The results in Appendix Table D.2 indicate that the results are also robust to these changes.

I also replicated the analyses in several different models where I try alternative empirical specifications and also test whether the results hold after excluding certain observations which might be misleadingly driving the results.

In order to check whether the results hold if I use a different statistical method, I replicated the analysis using a within-effects panel regression. The results I report in Appendix Table D.3 are similar to the original model and still support the main findings.

Next, I remove those provinces that were not yet established in the pre-1877 period, Beirut *(Beyrut)*, Bitlis, Kosovo *(Kosova)*, Mamüretülaziz, Mosul *(Musul)* and Van. Therefore, I replicate the analysis with the subset of the provinces that exist for all three periods (1868-1876, 1877-1896 and 1897-1910). The results in Appendix Table D.4 reveal that the findings are still the same.

I also remove the three provinces that were most affected by the severe famine that happened in Central Anatolia between 1873 and 1875 since the tax revenues of the state can be lower for

these provinces during the course of the famine and these provinces have higher percentages of Turkish population than most other provinces. This could be driving the results because the low tax revenues that may have happened during the famine make the changes before and after the war of 1877-8 higher than it would othewise have been. These three provinces are Kastamonu, Ankara and Konya (Quataert 1968). The results I report in Appendix Table D.5 are robust to the exclusion of these three provinces.

To account for the possibility that these results may be driven by provinces on which these wars were fought, I exclude these provinces from the analysis. The 1877-78 war with Russia was fought on Ottoman provinces of Edirne, Erzurum and Trabzon, while the 1897 war was not fought deep within Ottoman territory. The results remain very similar when these three provinces are excluded from the analysis, as I report in Appendix Table D.6. The most noteworthy change is that the coefficient for the interaction effect between ELF measure and Post-1877 Dummy declines in magnitude to -1.652. This coefficient is not estimated to be statistically distinguishable from 0 anymore at the 0.05 level of significance, but only is significant at the 0.1 level.

The final robustness check I conduct in this chapter is excluding the observations before the year 1871. This is the year when the second Provincial Law in the empire reorganized the bureaucratic and administrative structures within the provinces (Kırmızı 2010). I start the analyses in the year 1871 in order to account for the possibility that the administrative structures before 1871 might be driving the results. I report the results of this analysis in Appendix Table D.7, which are still in line with the expectations. The wartime increases in fiscal revenues are lower in provinces with higher diversity.

6.1.8 Alternative Explanations

In this section I evaluate several alternative explanations based on factors that the literature points out as determinants of the state's decision to invest in the fiscal capacity of some regions, or why it can successfully increase fiscal capacity during wartime. I demonstrate that there is no sufficient empirical evidence for some of these possible explanations and that the main results are

robust after accounting for these factors. For the remaining explanations that cannot be empirically tested, I discuss the reasons why they are not plausible in this context.

Ethnic Nationalism

One alternative explanation and rival mechanism one can worry about is ethnic nationalism. Taxpayers of Turkish ethnic identity might have been more likely to comply with higher taxes during war out of nationalist feelings and minorities may not have been willing to fund the state for such purposes due to the lack of such feelings, or negative feelings. Even though ethnic nationalism cannot explain why we see higher increase in more homogeneous places (as there is more increase in homogeneous provinces even after controlling for the percentage of Turks), it can be a rival mechanism to explain the higher wartime increases fiscal capacity where the percentage of Turks were higher.

Below, I argue that ethnic nationalism is unlikely to be the factor that causes these results because ethnic nationalism did not exist during this period among the Turks and many other minorities in the empire. Nevertheless, I also conduct an empirical analyses to demonstrate that even for the most likely case when ethnic nationalism can prevent tax extraction from a certain group, accounting for possible nationalism does not change the results. I focus on the Greek minority during the 1897 Greco-Ottoman war. The main results hold even after controlling for the proportion of the Greek minorities in each province.

Turkish nationalism is an ideology that first emerged among the elite and later diffused in a top-down manner to the larger population in a nation-building project during the Turkish Republic (Bayar 2014). It was not a prevalent ideology neither among the elite nor among the masses during the period that this study focuses on and it is extremely unlikely that the common people even had any ethnic consciousness. In fact, observers of the period write that even the meaning of the word 'Turk' was unclear and was sometimes used as an insult (Creasy 1854, p.7; see also Kushner 1977, pp.20-21).

The lack of ethnic consciousness among Turkish speaking Muslims in the empire also seems to be confirmed by a British officer who had travelled to the Ottoman Empire and wrote in 1908: "But if you say to a Mohammedan in Turkey 'are you a Turk?' he is offended, and probably answers, I am 'Osmanli' [Ottoman], or the Turkish equivalent of these words. An Osmanli Turk, if he says a man is a Turk, would mean that he is a lout or clodhopper" (Woods 1908, p.163; cf. Kushner 1977).

We can also observe the repercussions of this phenomenon in Turkish literature. Famous author, intellectual and politician in the early Turkish Republic, Yakup Kadri Karaosmanoğlu criticizes the Anatolian villagers of Turkish ethnicity for their lack of support for the Turkish War of Independence (1918-1923) and their ignorance of Turkish nationalism (Karaömerlioğlu 2002) and the following dialogue between a nationalist former army officer and a villager from his famous novel Stranger (Yaban) offers an important clue to how the concept of ethnicity and the ideology of nationalism were alien to an ordinary Ottoman citizen of Turkish ethnicity even as late as the Turkish War of Independence:

> Villager (V): "I know sir, you are one of them."
>
> Officer (O): "Who are they?"
>
> V: "Those who are on the side of Kemal Paşa."[9]
>
> O: "How can someone be Turkish and not be on the side of Kemal Paşa?"
>
> V: "We are not Turkish, sir."
>
> O: "What are you then?"
>
> V: "We are Muslims, praise to God... Those live in Haymana." (Karaosmanoğlu 2003 [1932], pp. 152-3. Own translation).[10]

---

[9]Refers to Mustafa Kemal Atatürk, who was leading the Turkish War of Independence.

[10]The original text in Turkish is:

> "Biliyorum beyim, sen de onlardansın, emme."
> "Onlar kim?"
> "Aha, Kemal Paşa'dan yana olanlar..."
> "İnsan Türk olur da nasıl Kemal Paşa'dan yana olmaz?"
> "Biz Türk değiliz ki beyim."

Eugene Weber mentions two examples from the period before ethnic nationalism spread among the masses that is reminiscent of this passage from Karaosmanoğlu's novel. His example from the year 1706 in France is about Dominican Emmanuel Labat, who was "struck" that the residents of Marseille rejected that they were French (Weber 1976, p.98). Another example he mentions resonates well regarding the perceptions of belonging of people to countries, and how they perceive wars. This is a Peruvian novel "about Indian peasants who thought that a war with Chile was being waged against a general of that name, and were nonplussed when told that Chile was one country and Peru another to which they belonged." (p. 108). Similar to the examples in these contexts, it is unlikely that without any ethnic or nationalist consciousness, the Turkish-speaking people in the Ottoman Empire would perceive wars as being fought against a country of another ethnic group by their country representing their ethnic group.

Another alternative explanation here can be a sort of 'Muslim nationalism' where taxpayers were motivated by waging war against countries with non-Muslim populations by their Muslim country. However, the evidence from local-level tax revenues I presented above in addition to the evidence I will discuss in Section 6.3 reveal that the Ottomans had difficulty in building fiscal capacity in areas where non-Turkish Muslim populations such as Albanians, Arabs or Kurds lived.

Regarding non-Turkish ethnic groups, in many diverse provinces in the Ottoman Empire some minority populations had not yet embraced any nationalist ideology and some minority nationalisms were yet nonexistent. Reinkowski dates the start of a prevalent and effective nationalism in the Arab lands, Turkey, and Albania to the early twentieth century, or even later (Reinkowski 2017, p.16). We can see similar arguments in the works of other historians on Kurdish nationalism (Özoğlu 2001), and also on Arab nationalism whether on Ottoman lands (Haddad 1994), or along the Ottoman-Persian border (Ahmadi 2021).

Arabs and Kurds comprised a vast majority (or almost all) of the population in the provinces of Aleppo *(Halep)*, Baghdad *(Bağdat)*, Basra, Beirut *(Beyrut)*, Hejaz *(Hicaz)*, Syria *(Suriye)*, Tripoli

---

"Ya nesiniz?"
"Biz Islam'iz Elhamdülillah... O senin dediklerin Haymana'da yaşarlar."

of Libya *(Trablusgarp)* and Yemen. In addition to this, they constitute a significant portion of the populations in the provinces of Bitlis, Diyarbakır, Erzurum, Mamüretülaziz and Van. With substantial populations of these two ethnic groups in so many provinces and lack of ethnic/nationalist consciousness among them, it is unlikely that the main results in this chapter occur because these groups do not comply with taxation during wartime because of their nationalist ideologies or feelings.

In any event, in order to systematically account for ethnic nationalism in the empirical analyses, I calculated the proportion of Greeks in each province and added it as a control variable to see if it changes the results for the 1897 Greco-Ottoman War. This is an ideal case to study ethnic nationalism as an alternative mechanism because the Ottoman state fought against the Greek nation state in this case and substantial numbers of the members of the Greek minority lived in many Ottoman provinces. Furthermore, Greek nationalism was more salient than nationalism among many other minority groups in the empire and the Greek minority in some regions of the empire had fought a successful War of Independence against the empire as early as the 1820s, establishing the Greek nation state. If ethnic nationalism can ever alter the results on diversity I identify above, this is the most likely case when it can do so.

To account for Greek ethnic nationalism during the 1897 Greco-Ottoman War, I interact the proportion of the Greek population in each province with the post-1897 dummy. This way, I control for the percentage of Greeks in each province. The results I report in Table 6.12 reveal that controlling for the percentage of the Greek population does not change the results for the effect of diversity on the changes in fiscal revenue during the 1897 War.

Risk of Uprising

The second alternative mechanism can be related to the first one, ethnic nationalism. If in more diverse areas higher taxes are more likely to trigger the minorities' nationalist feelings and make the minorities more likely to start or participate in an uprising, especially if this could create a risk of secession, then the state may have preferred to increase the tax burden on less diverse areas.

Table 6.12: Generalized Difference-in-Differences Analysis of Wartime Increases in Fiscal Revenues, Controlling for the Proportion of the Greek Population

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	21.900***
	(1.252)
Percent Turkish	10.522***
	(1.188)
Post-1877 Dummy	6.917**
	(3.181)
Post-1897 Dummy	6.820**
	(2.992)
Percent Greek * Post-1897 Dummy	−2.073**
	(0.858)
ELF * Post-1877 Dummy	−2.023**
	(0.831)
ELF * Post-1897 Dummy	−2.411***
	(0.794)
Percent Turkish * Post-1877 Dummy	2.854***
	(0.823)
Percent Turkish * Post-1897 Dummy	4.158***
	(0.809)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	761
R^2	0.897
Adjusted R^2	0.886

Note: OLS Regression. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Again, this logic cannot explain why there are higher increases in fiscal revenues in more homogeneous provinces. In fact, the risk of uprising should be higher in more homogeneous provinces because the collective action problem can be more easily overcome among more homogeneous groups. Still, it is worth discussing this alternative mechanism as a potential rival for the Ethnic Similarity (Percent Turkish) variable.

In order for risk of uprising to be a concern here, we should be able to observe two factors. First, there should be widespread nationalist feelings among the Ottoman populations. In the previous

section I argued that this is not true for most of the empire, except for some limited populations in the Balkans. The second factor that should be present for risk of uprising to have shaped the state's patterns of investment in different localities and fiscal capacity is that the Ottoman State was aware of the risk of uprising and pursued investment strategies in response to these perceived risks. It is difficult to argue that the Ottomans were aware of Nationalism as a phenomenon and understood how it could trigger uprisings. Davison (1977) writes that there is no evidence that the Ottomans had recognized and understood the strong nationalist aspect of the uprisings of the nineteenth century.[11] To them, these uprisings probably were not any different than any other uprising that had occurred in the previous centuries.

There is also not much evidence that tax-related uprisings were more likely to happen among the minority populations. In fact, discussing the tax revolts across the Ottoman Empire that happened before the 1908 revolution, Aytekin (2013, p.323) argues that there were more tax-related uprisings in Central and Eastern Anatolia than other regions in the empire and that most rebels were Muslims. Central Anatolia is predominantly Sunni Muslim Turkish.

One straightforward explanation for this pattern can be that there were more uprising among the Muslim populations because their tax burdens increased heavily while minorities were less likely rebel because their tax burdens did not increase, and this was a consequence of a calculation by the Ottoman state, which aimed to minimize nationalist uprisings in regions with more minority populations as they would be likely to break up from the Ottoman Empire, and this risk was smaller for Muslim Turks, and putting the tax burdens on Muslim Turks rather than other groups.

I argue that such a perspective cannot be entirely accurate. It cannot explain why we see uprisings in both Central Anatolia, where Sunni Muslim Turks formed a vast majority of the populations and Eastern Anatolia, where they were a minority. In addition to this, this perspective cannot explain why in other regions with Sunni Muslim Turkish majorities such as Western Anatolia (the provinces of Aydın and Hüdavendigar), two of the provinces with the highest increases in per

---

[11]The nationalist uprisings in the nineteenth century include Serbian, Greek, Montenegrin, and Bulgarian uprisings.

capita revenues during the two wars that I examine, the extent of uprisings were behind Central and Eastern Anatolia.

Initial Fiscal Capacity

Another alternative explanation can be about scales, or ceiling effects. It might be easier to achieve higher increases in regions where the per capita fiscal revenues were lower to start with and in places where the per capita revenues are already high it might be more difficult to increase them because they are already very high.

This mechanism is unlikely to be the reason why we see higher increases in fiscal revenues in less diverse regions because in the Ottoman empire the per capita fiscal revenues were very low compared to many other European countries (Karaman and Pamuk 2010), and the European states could still increase their fiscal revenues (Mitchell 2007). This was true for all the provinces in the empire. Therefore, there was substantial room for improvement of the state's fiscal revenues regardless of the initial levels.

In Figure 6.4 I present the average per capita revenue and the ELF measure by period for each province. In Figure 6.5 I report the same for the Percent Turkish variable. In order to make comparisons easier I present each measure in a binary manner, above 0.5 and below 0.5 (remember that they are both continuous). The descriptive patterns in these figures suggest that the increases in fiscal capacity are unlikely to be driven by provinces that start with low fiscal capacities and are less diverse.

Urbanization

Another alternative mechanism to worry about can be levels of urbanization (Tilly 1992; Abramson 2017) since it may have determined the levels of state strength. Regarding fiscal capacity, higher rates of urbanization is expected to influence tax revenues overall because it is easier to monitor and assess urban tax bases compared to rural and agricultural tax bases and urban eco-

Figure 6.4: Tax Revenue Per Capita for each Period, by Province and ELF
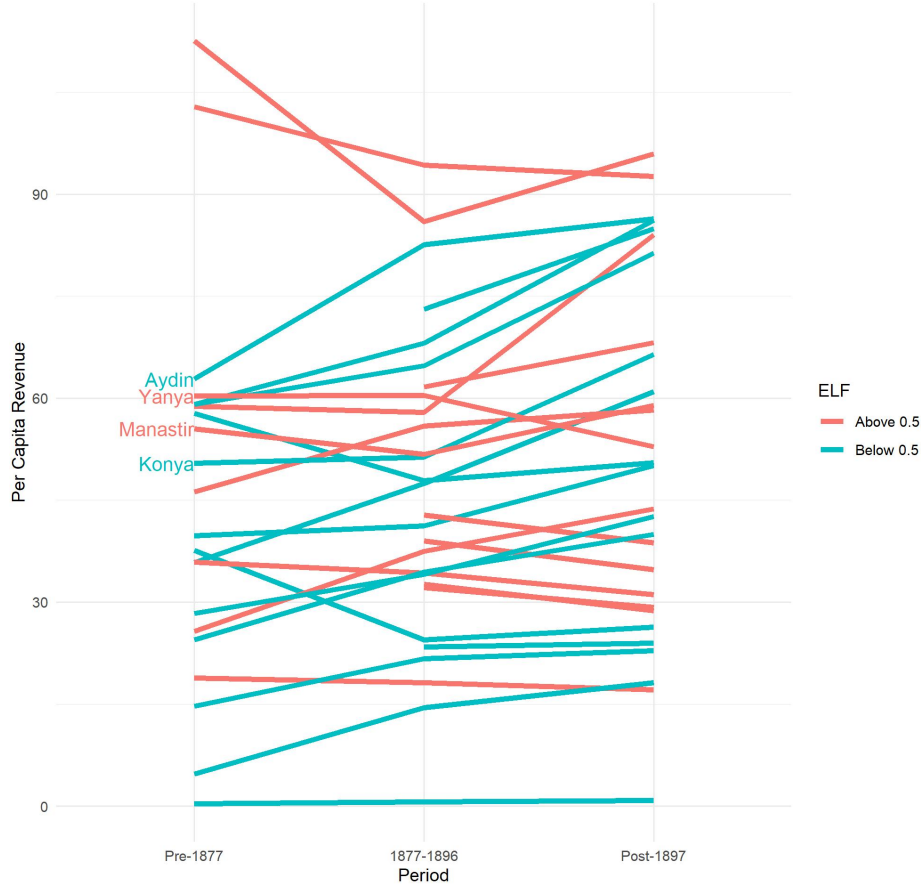
nomic transactions are more monetized (Karaman and Pamuk 2013) or because urban populations

have preferences that make them more acceptant of taxation (Andersson 2018).

The ideal way to test this alternative mechanism would be to collect time-variant data on the

urbanization rates in each province and control for this covariate in the analysis. Unfortunately, I

have not been able to identify temporal data for this variable and the data I identified so far does

not include measures for many of the provinces (there is data for only sixteen provinces out of

twenty-nine), which forces me to evaluate this alternative mechanism in a descriptive way with the

data I have.

Urbanization should not be a factor to be worried about in the context of this study unless it is

negatively correlated with the ELF and positively correlated with the ES measures. Using Cuinet's

(1890-95) work that reports the population of certain Ottoman urban centers, I constructed urban-

Figure 6.5: Tax Revenue Per Capita for each Period, by Province and ES
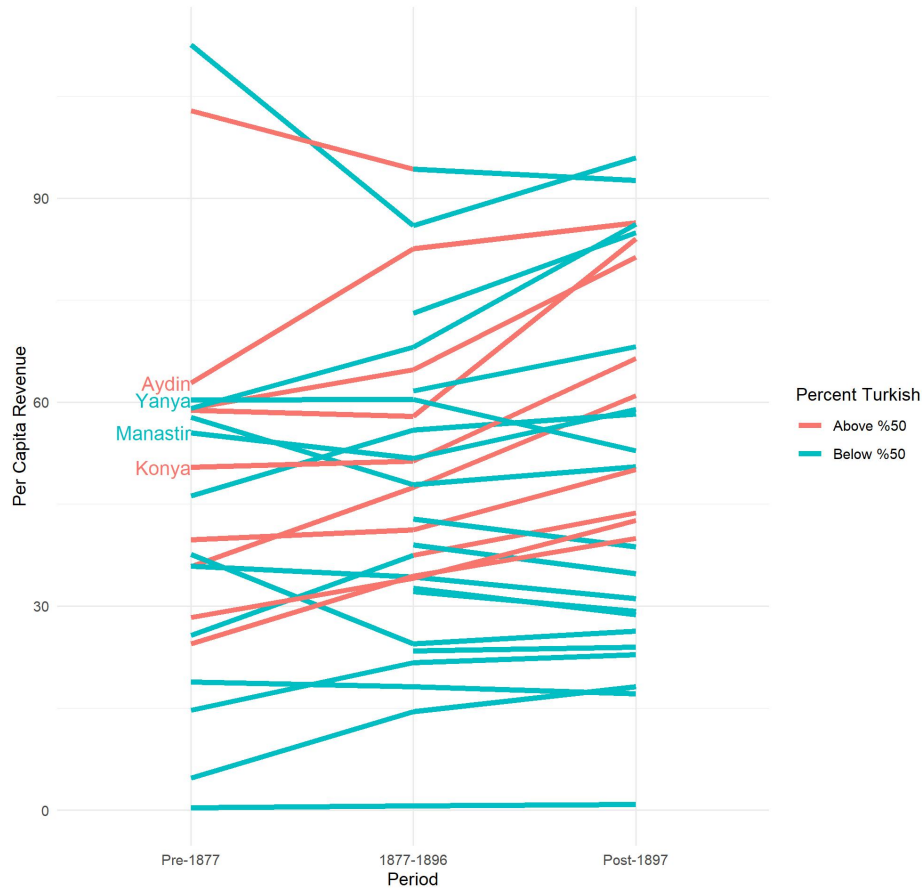
izization measures for each province for which Cuinet provides data. There is positive correlation (Pearson $r = 0.57$) between the ELF measure and the urbanization, while a slight negative correlation (Pearson $r = -0.18$) between the ES measure and urbanization. Neither of these relationships indicate that urbanization is correlated with ELF and ES in a way that can explain the results. The scatterplots in Figure 6.6 plot the relationships between these variables.

Economic Output/GDP

GDP can be another alternative explanation worth considering since there has been empirical evidence in the literature that higher GDP enables higher tax revenue per capita (Besley and Persson 2011). The concern here can be that the wartime increases happen in provinces with higher GDP because there is more economic activity to tax in these provinces. However, GDP should not be

111

Figure 6.6: Urbanization and Ethnic Composition
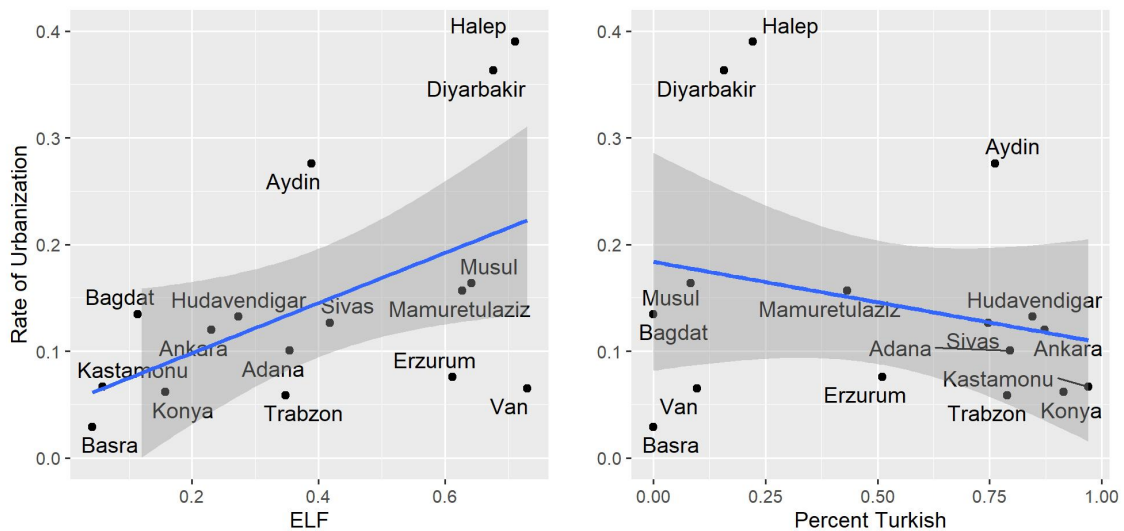
a problem unless it is negatively correlated with ELF and positively correlated with ES measures. Province-level GDP estimates are available in the 1897 Statistics Book published by the Ottoman State (Güran 1997). The scatterplots I present in Figure 6.7 below indicate that GDP per capita has a slight positive correlation with ELF at the province level, and is negatively correlated with ES. Neither of these factors, at least in the year 1897, correlate in a way with GDP per capita to indicate that GDP per capita can be the factor that can cause the outcome.

Next, I go on to add this measure of GDP as an interactive control variable in the generalized difference-in-differences model. Even though a time-variant GDP variable would be ideal to control for GDP levels in the analysis, due to the lack of such data the best I can do is to use the time-invariant GDP measure. The results in Table 6.13 demonstrate that the main results are robust to controlling for the GDP. Note that the model drops the main effect for the GDP variable because it is time-invariant.

Technology Transfer

A final alternative mechanism to consider can be the easier transferability of technology and innovations in places with homogeneous populations and across places with groups of similar ethnic

112

Figure 6.7: GDP per Male Capita and Ethnic Composition
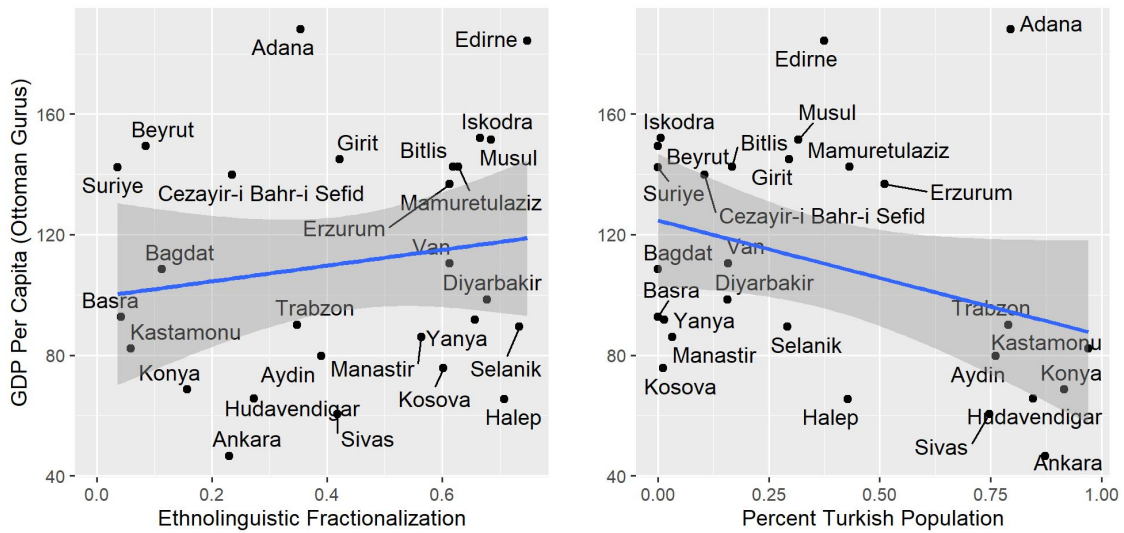
identity. According to Tilly, "In a homogeneous, connected population, an administrative innova-

tion installed and tested in one region had a reasonable chance of working elsewhere, and officials

could easily transfer their knowledge from one locality to another" (1992, p.100). Nevertheless, I

do not think that we can consider this mechanism as separate from legibility. First and foremost, it

is impossible for the state to know the ethnic composition of the population in certain regions with-

out rendering these populations legible. Only after this it can decide what technology to transfer

here, according to the technologies that were successful under similar ethnic compositions. Fur-

thermore, merely learning about the ethnic composition is not the only problem. The state would

also need to learn about the economic activities, such as the types of products that are grown so

that it can transfer new taxation technology here. This, again, requires that the legibility problem

be solved.

Table 6.13: Generalized Difference-in-Differences Analysis of Wartime Increases in Fiscal Revenues, Controlling for the GDP Levels

	Dependent Variable:
	Revenue Per Male Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	22.320***
	(1.361)
Percent Turkish	10.708***
	(1.221)
Post-1877 Dummy	3.605
	(4.161)
Post-1897 Dummy	6.595
	(4.126)
ELF * Post-1877 Dummy	−1.796**
	(0.881)
ELF * Post-1897 Dummy	−2.792***
	(0.839)
Percent Turkish * Post-1877 Dummy	3.007***
	(0.874)
Percent Turkish * Post-1897 Dummy	3.566***
	(0.869)
GDP (million guruş) * Post-1877 Dummy	0.035
	(0.031)
GDP (million guruş) * Post-1897 Dummy	0.021
	(0.035)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	736
R^2	0.893
Adjusted R^2	0.880

Note: OLS Regression. Standard errors in parentheses.   *p<0.1; **p<0.05; ***p<0.01

6.2 Evidence from Individual-Level Taxation Data

6.2.1 Data and Variables

In this section, I use data from Income and Property Registers (Temettüat Defterleri) which were compiled during the property censuses conducted in the 1840s. These registers contain information about the taxpayer's name, occupation, properties, the amount of taxes that the taxpayer was

assessed to be liable for, at the individual/household level.[12] The religious identity of the individuals can be easily inferred from the names of the individuals, and in some registers the religious identity of the individual is recorded.

I use an already-transliterated register for Monastir's central kaza (third-level administrative unit), from Alimoski (2005). The town center that served as Monastir's central kaza was called Manastır in Turkish by the Ottomans. It is within the national boundaries of North Macedonia today, and is now known by the name Bitola.

This register contains information about the identity of the taxpayer (Muslim, Christian, Jewish, Muslim Roma, Christian Roma). Unfortunately there is not much information that can allow one to infer the ethnic identity of the Muslim and Christian taxpayers, except for the Roma. This is why I conduct the analysis based on the religious identity of the taxpayer. I exclude the Roma from the sample because they are often much poorer compared to non-Roma similar to with anywhere else in Europe and Middle East then, as they are today. Based on the demographic information of the area, the Muslims living here are likely to consist of Albanians and Turks.

I use the identity as the independent variable, constructing a dummy variable that equals 1 if the taxpayer is Muslim, and 0 if the taxpayer is Jewish or Christian. In alternative models, I use Muslim as the baseline category and have separate dummies for Jewish and Christian taxpayers.

The analysis with individual-level taxation data here serves the following purposes. First, even though I attempted to alleviate concerns about ecological inference problems in the province-level data by replicating the analyses with sancak-level data in Section 6.1, all the data I presented in that section was at the aggregate level. In addition to this, the sancak-level data contains too many missing values and the analysis can be conducted with only a small subset of all the sancaks. In this section, I aim to alleviate these concerns by present some evidence from individual-level taxation data from the Ottoman Empire.

Another purpose of this analysis is to alleviate the concerns regarding the results in Section 6.1 where I found more heterogeneous provinces to be paying more capita taxes overall. I argued

---

[12]Women are not counted in these censuses unless they are the head of the household (which is the case only when they were widowed) and were the taxpayers. Children are also not counted.

that overall levels of fiscal capacity can be misleading at the aggregate-level data because the overall levels of fiscal revenues in different provinces are likely to be driven by the overall levels of economic output. Here, by looking at taxation patterns within restricted geographical area, I can hold many other factors constant, including the potential for economic output.

Obviously, there are downsides to this analysis. The data being from a very restricted geographical area can raise concerns about the generalizability of the results. The patterns can be specific to this sample. I also cannot directly test the taxation patterns of different ethnolinguistic groups, but have to focus on taxation patterns according to religion. Another weakness of an analysis using this data is that the data from the 1840s, while the aggregate-level data I used in Section 6.1 starts in 1868. At the end, this is the best I can do given the circumstances as the options regarding individual-level data are limited here and this is the best data that I could find.

6.2.2 Results

The results I present using OLS Regressions in Table 6.14 indicate that Muslim residents of the town of Monastir were assessed to be responsible to pay significantly more taxes compared to its non-Muslim residents. The registers contain information about the neighborhood that the taxpayer resides, and also their occupation. Even though neighborhood and occupation should ideally be controlled since they can influence the income and wealth levels of individuals and affect how much taxes they were responsible for, in this case there is significant segregation into different neighborhoods and occupations by religious identity. This is why adding occupation and neighborhood as covariates is not possible in the analyses I present here.

In Model 1, I regress total tax obligation on Muslim Dummy. On average, Muslims are responsible for paying 145.2 guruşes higher taxes compared to non-Muslims. The mean taxes that non-Muslims are responsible for paying is 412.4 guruşes. This indicates a drastic differece between Muslims and non-Muslims. The average Muslim pays %35.2 more taxes compared to an average non-Muslim in Monastir's central kaza.

Table 6.14: Analysis of Individual-Level Tax Payments by Identity in Monastir's Central Kaza

	Dependent Variable:					
	Total Tax Assessment (in gurus)					
	(1)	(2)	(3)	(4)	(5)	(6)
Muslim Dummy (vs. Non-Muslim)	145.2***		199.4***		99.1***	
	(34.0)		(36.4)		(27.3)	
Christian Dummy (vs. Muslim)		−122.9***		−175.0***		−73.2***
		(35.4)		(37.8)		(28.3)
Jewish Dummy (vs. Muslim)		−267.8***		−330.3***		−242.2***
		(63.2)		(65.9)		(48.6)
Constant	412.4***	557.6***	423.8***	623.2***	413.3***	512.4***
	(22.8)	(25.3)	(23.8)	(27.5)	(17.1)	(21.3)
Observations	4,462	4,462	4,166	4,166	3,820	3,820
R^2	0.004	0.01	0.01	0.01	0.003	0.01
Adjusted R^2	0.004	0.005	0.01	0.01	0.003	0.01

*Notes: OLS regression. Standard errors in parentheses. Models 1 and 2 include all the observations. Models 3 and 4 exclude observations whose occupations indicate they may not have regular incomes. Models 5 and 6 exludes these groups in addition to state officials. *p<0.1; **p<0.05; ***p<0.01.*

In order to check whether the same patterns exist between Muslims and Jews as well as between Muslims and Christians, I treat Muslim as the baseline category in Model 2 and regress tax assessment on separate Christian and Jewish dummies. Muslims are responsible for paying higher taxes compared to Christians and also compared to Jews.

In Model 3 of Table 6.14 I replicate Model 1 by excluding the observations whose occupations are recorded as students, unemployed, receiving charity, fugitive and those where the occupation could not be read, and in Model 4 I replicate Model 2 with this subsample. Those recorded under these categories may not have regular incomes, and may be paying lower taxes. The main results do not change. If anything, the differences between Muslims and the other groups are larger for this subsample.

In the last two models I additionally exclude state officials from the sample because their income and property might be more visible to the state and they can be taxed more easily. [13] Additionally, the sample includes several military and bureaucratic elites, whose incomes and might

---

[13] In Appendix B I present those occupational categories that I coded as state officials.

be higher than the average taxpayer in this sample, and who may have accumulated properties due to the privileges high-strata Ottoman Elite enjoyed until these privileges were repealed in 1839 (Cansunar and Kuran 2019). These elites are more likely to be Muslim. The results are again similar. Muslims, on average, pay higher taxes compared to non-Muslims according to Model 5 of Table 6.14, even though the estimate declines to 99.1. In Model 6, the estimates for the Christian Dummy and the Jewish Dummy are also smaller, but the results overall indicate the same patterns. Christians and Jews in Monastir's central kaza pay lower taxes compared to Muslims.

One concern with this analysis is that Muslims could be assessed to be responsible for higher taxes as they had higher incomes. This is possible, but from what we know about the Ottoman Empire during this era is that Non-Muslims, and especially Christians were economically better-off than the Muslims. Non-Muslims were more likely to be active in trade (Kuran 2004), and in the agricultural sector while Muslims were more often growing traditional grains, non-Muslims were growing valuable cash crops for the export markets (Issawi 1982, p.264). Therefore, if there is any difference in wealth and income across Muslims and non-Muslims that can drive different levels of tax assessment across these groups, it should normally be in the opposite direction.

In addition to this, until the early twentieth century, Muslims had to serve in the military but non-Muslims had to pay a poll tax instead of serving. This should normally mean that non-Muslims are expected to pay higher taxes because of two reasons. First, serving in the military for years, and sometimes a decade (Zürcher 1998), should create economic disadvantages for Muslims because they will lack experience and connections that can help them earning higher incomes after returning from military service. Second, the poll tax that the non-Muslims pay and the Muslims do not pay should straightforwardly increase the tax burdens of the non-Muslims compared to Muslims.

Restricting the analysis to such a small geographical area, to only one kaza, obviously does not alleviate all the concerns that I try to alleviate regarding the aggregate-level taxation data. Furthermore, it is still possible that this specific kaza had patterns of inter-religious group wealth and income distribution that is different than the patterns in the Ottoman Empire in general. Neverthe-

less, the results I discussed here are in line with the expectations I put forward and the empirical analyses I presented in the preceding sections.

6.3 Evidence from Archival Documents and Secondary Sources

In the preceding sections of this chapter, I used a local-revenue dataset to demonstrate that the Ottoman state was less likely to increase its fiscal revenues where the populations were more diverse. In this section, I present and discuss evidence from additional sources, including correspondences and reports in the Ottoman Archives and secondary sources as additional evidence in line with the main results I obtained with the local-level fiscal revenue data. In discussing these materials in this chapter, I aim to provide evidence, that builds on the quantitative findings I have presented so far, and in support of the mechanisms I offer. I aim to demonstrate that there is evidence that diversity was a key factor in limiting the state's capacity.

I discuss the archival evidence and evidence from secondary sources under three categories in the next three subsections. In the first subsection, I discuss evidence from archival sources that demonstrate how diversity made it more difficult for the Ottoman state to increase their administrative and fiscal capacity. In the second, I provide evidence that diversity of the population made the intermediaries stronger, increasing their bargaining power. In the last subsection, I demonstrate that the difficulties that the central state encountered in tax collection also existed in another configuration of the state's power and capacity, in military conscription.

6.3.1 How Diversity Undermines Tax Collection

The Ottoman state faced significant difficulties in projecting their rule towards the diverse regions of the empire due to linguistic and religious differences between the state's agents and the population. The problems of illegibility emanating from these differences made the Ottoman state struggle in taxing the minority populations. Even as late as the 1860s, there were parts of the Arab Provinces where the state could not raise any taxes (BOA.A.}MKT.UM. 532/59; BOA.A}.MKT.MHM. 349/68). The differences between the state's agents and the population's

identity undermined tax collection because they contributed to the difficulties that the state encountered in appointing better-skilled bureaucrats and thus failing to increase bureaucratic capacity. They also undermined tax collection because they negatively contributed to the unsuccessful attempts (or lack of attempts) of the state in counting the populations. An ideal system of taxation would require that the state count the population at regular intervals, and also update the assessment of taxes every year. When it could not count the populations, update its knowledge about their economic activities at regular intervals, and reassess the taxes that would be levied on the populations every year, the Ottoman state often had to rely on not-up-to-date population counts and based its tax assessment on the taxes that were paid in the previous years (BOA.İ.MVL. 279/10885).

Linguistic problems posed important obstacles towards the central government's attempts to hiring more skilled tax administrators and taxing the populations. One example where this is evident is Interior Minister Memduh Paşa's letter to the office of the Prime Minister in 1902, about hiring new tax collectors in the province of Mediterranean Islands *(Cezayir-i Bahr-i Sefid)* (BOA.BEO. 1968/14754). The province, consisting of the Greek islands situated between today's mainland Greece and Turkey, had a Greek-speaking Orthodox Christian majority.

In this letter, Interior Minister Memduh Paşa responds to a telegram from the prime minister *(sadrazam)*, which had demanded that in the province of Mediterranean Islands tax collectors should keep their books in Turkish and not in Greek. The same telegram also apparently communicated that assigning tax collectors from other provinces to this province was out of question because of the high salaries that would need to be paid.

The Interior Minister explains to the Prime Minister that in some parts of the province hiring new Turkish-speaking tax collectors is impossible because of the lack of Turkish speakers in the province who are literate. He goes on to add that the previous tax collectors who already knew the people in the area and who were skilled in the techniques of collecting taxes should be allowed to resume their work even if they do not speak Turkish and cannot keep their books in Turkish. He also asks for the same for the tax collectors who speak Turkish, but are not well-trained enough for bookkeeping in the new techniques. Finally, he asks the prime minister to continue allowing the

practice of bookkeeping in Greek, presumably as many of these desired tax collectors speak only Greek but not Turkish.

This describes a phenomenon where linguistic differences are likely to constrain the state's fiscal revenues. Ideally, the tax collectors should speak the language of the administration (Turkish), the language of the tax base (Greek), know the area and the people to efficiently tax them, are skillful tax collectors, and keep their books in Turkish so that the books can be inspected by any agents sent by the central state.

From this letter we understand that there are few Turkish speakers in the province who can fulfil this duty. This results in the minister asking for two changes from their initial demands. The first is that the Turkish-speaking tax collectors keep their job even if they are not well-trained enough for bookkeeping in the new techniques. This would result in inefficiencies in increasing tax revenues because such less advanced bookkeeping is likely to make inspection from above more difficult. With lack of inspection, corruption is easier and the revenue that could have ended up in state treasury is likely to have stayed in the pockets of the taxpayer, and ended up in the agents' hands.

The second change that the minister asks should result in a similar problem for the state's fiscal capacity. The lack of skilled tax collectors who knew the area and people leads the minister to allow the Greek-speaking tax collectors who do not speak any Turkish and allow them to keep their books in Greek. This, again, is suboptimal for higher capacity because such books can only be inspected by superiors who speak Greek. Such a limit on the possibility of inspection of the books will make it more likely that any abuses such as embezzlement by the tax collector will go unnoticed, unless the state can invest more money into sending Greek-speaking officials for inspection. The problem of bookkeeping in linguistically diverse regions is a recurring theme in Ottoman taxation, where the ability to keep books in Turkish was often a requirement to be hired as tax collector (BOA.İ.MVL. 560/15166).

Another letter from the year 1907 provides another example where efficient tax collection and administration ran into the problem of linguistic diversity. The administrator who penned this

letter, which was sent to the head treasurer of the province of Manastir from its district Kozani (*Kozan*) in today's Greece, complains that the tax office scribe who was assigned to the district and was responsible for the areas outside the city did not speak the local language (BOA.TFR.I.MN. 138/13707).[14] The administrator writes that this is a problem as there are many villages where only Greek is spoken and it would prevent this person from achieving the improvements *(tadilat)* in these villages. Hence, the letter asks that, in addition to the tax collectors who were required to speak local languages according to the regulations, all the tax office clerks who are assigned to different townships and villages in the district speak the local language language there, in addition to Greek. This implies that officials even at such lower levels in the administration are expected to have very high skills due to linguistic diversity. They need to speak one local language in addition to Greek. From the letter, it is not possible to learn whether knowledge of Turkish was necessary for this job, although it probably was necessary because the state needed and often preferred bureaucrats who could keep records and communicate in Turkish. If so, the linguistic diversity in this case required scribes who spoke and could perform their tasks in at least three languages. This makes it more difficult to find such people and even if they can be found, it should be very costly to hire them. These factors are likely to undermine fiscal capacity.

In a famous report *(layiha)* composed in the year 1880 by the governor of the Baghdad province, Abdurrahman Paşa, we can see how high-ranking administrators of the empire complained about the complications that diversity creates in ruling and taxing the populations (BOA.Y.EE.7/12). Abdurrahman Paşa's report is part of an attempt to acquire information about and implement reforms in the Baghdad province, which roughly covers today's modern-day Iraq.[15] This report summarizes all the reports that Governor Abdurrahman Paşa had received from district (sancak) and subdistrict (kaza) governors. He writes that the most urgent problem in Ottoman Iraq was the lack of security and order, and he blames the "very different cultures, sects and customs" (Çetinsaya 2006, p.25)

---

[14]This local language is not specified, but from the rest of the telegram it is understood that the language in question is Greek.

[15]Later, Mosul and Basra provinces would be separated from the Baghdad province.

122

of the people in the region for this problem. He remarks that the only way to collect taxes and conscript soldiers was by force, which magnifies the problem of security and order.

After the Ottoman government became insolvent and defaulted on its debt in 1875 and 1876, the Ottoman Public Debt Administration (OPDA; *Düyun-ı Umumiye*) was founded in 1881. This European-controlled institution collected some sources of state revenues in order to pay back the debt. The OPDA at some point in its existence had 5,000 employees, including those who were involved in the collection of these tax revenues (Quataert 2005, p.72). Even though the OPDA was overall successful, an example about tobacco tithe collection from the province of Shkoder *(İşkodra)* indicates that its functioning did not go that smoothly in more diverse areas.[16] Tobacco tithe was one of the revenue items that were collected by the OPDA. In a telegram from 1898 (BOA.DH.MKT. 2091/36), it is reported that the level of revenues from tobacco tithe were very low *(dun bir raddede)* because those who were responsible for the assessment and collection of the tithe could not adequately keep up with the task *(layıkıyla takip-i maslahat olunamamasından)*. In this case, an otherwise well-functioning institution has trouble collecting taxes from this very diverse province years seventeen years after it was founded.

Often not having sufficient administrative capacity to rule and tax the minority populations the Ottomans had to rely on military force to collect taxes in many places. Anderson (1986, p.66) notes that in the early nineteenth century Tunisia the Ottomans had to send a military expedition twice a year to collect taxes from the rural populations, and Talhamy (2011, p.34) writes that troops were necessary to collect taxes and recruit conscripts from the Arab Alawite population in the Levant. Similarly, Özbek (2015, pp.191-2) writes that in parts of the province of Aleppo where the tribal populations lived it was impossible to send tax collectors if one or two soldiers did not accompany them, while in the districts of Najd *(Neced)*, Ammara *(Amare)* and Muntafiq *(Müntefik)* in Iraq where most of the population were nomadic tribes the helplessness of the state agents responsible for tax collection made it necessary to send a platoon of soldiers *(müfreze-i askeriye)* for tax collection.

---

[16]See Birdal (2010) for an extensive study of Ottoman debts and the OPDA.

Having to send military or other security forces to collect taxes is a symptom of the insufficient control and bureaucratic or administrative capacity of the empire in these provinces. This likely is not ideal for increases in fiscal capacity because rather than civil servants who are specialized in tax assessment and collection, here the tax collection duties are delegated to the military, whose performance may be worse than professional tax collectors. Even if the military accompanied the tax collectors, it makes tax collection much costlier than only sending tax collectors.

The reliance on local intermediaries also allowed them to engage in frequent abuses of the system, undermining the state's fiscal revenues. In Nablus during mid-nineteenth century, the local intermediaries in the Nablus Advisory Council challenged the state's attempts to increase its revenues from tax farming by sabotaging the tax farming contract bids made by outsiders and trying to keep the tax farm contracts to themselves (Doumani 1995, p.242). The abuses by the intermediaries also often put a heavy burden on the taxpayers, and the complaints and news of such abuses was difficult to convey to the state's agents, probably due to language problems. The intermediaries who spoke the local language were often the only link between the state and the locals. When they abused the taxpayers, there were no channels to complain to other authorities about such abuses. In one letter from the year 1862, the Mutasarrıf of Van and Hakkari (in today's South-Eastern Turkey) acknowledges the receipt of a letter from the center which orders him to conduct investigations about the abuses committed by the Christian community leaders and muhtars, where they are accused of collecting extra money for themselves from the taxpayers during the collection of taxes from them (BOA.A.MKT.UM. 532/59).[17]

Complaints about similar abuses by local intermediaries is in Muhammed Hilal Efendi's report about Yemen that I discussed in Chapter 4 (BOA.YEE. 58/33). He notes that one consequence of the state not possessing information about the names of the villages in the province of Yemen was that the collection of taxes was delegated to sheikhs, which according to him caused inconveniences *(uygunsuzluklar husule gelmektedir).* His proposed solution to these inconveniences was to conduct a population and property census, yet I have argued in Chapter 4 that this was a difficult

---

[17]Muhtars were the heads of the villages or neighborhoods.

task due to the inability of the state to count the populations. Large-scale corruption in agricultural tithe collection, including high-level bureaucrats in Yemen is reported in another document from the year 1887 (BOA.DH.MKT. 1417/5).

6.3.2   Intermediaries Stronger under Diversity

Above, I have described instances of how diversity caused problems in tax administration and tax collection. In this subsection I provide evidence from the correspondences and secondary literature on how diversity increased the premium on local intermediaries, making them indispensable to rule and tax the diverse populations, giving them higher bargaining power against the state and allowing them to resist state building.

Diverse states, especially empires, often relied on indirect rule where they ruled through local intermediaries (Afigbo 1972; Doyle 1986; Tilly 1992). Having higher bargaining power against the state, many intermediaries used this to their own advantage and undermined the state's capacity, depressing the state's tax revenues in areas under their influence and control.

The local intermediaries have informational and network advantages compared to an outsider who could be assigned from the center. They often had much better information about what was going on in the periphery than the central state and its agents. A telegram from the center to the Bitlis province in Eastern Anatolia from the year 1902 is an appropriate example to show how little information the state had about the peripheries and how critical a link intermediaries were between the population and the state (BOA.DH.TMIK.M. 132/42). From this telegram we learn that in the Armenian-populated villages of Muş sancak the tax collectors had engaged in abuses *(suistimal)* and coercion *(tazyikat)* during tax collection.

The center demands from the Bitlis province that they investigate these allegations of abuses and coercion. What is interesting here is that the center tells the Bitlis province that it had learned these complaints not directly through the villagers or the head of villages *(muhtars)*, but through the Armenian Patriarchate. The muhtars wrote the letters to the Patriarchate, not to the state. It is unlikely that muhtars, let alone the villagers, were able to write letters in Turkish to be able to

125

correspond with the province, or the center. It is not possible to learn whether the tax collectors involved in abuses and coercion are the central state's agents or agents of tax farmers. Of course, it may be that those involved in these abuses were state's agents and bureaucrats at the province center were aware of this, but took no action. Even if this is the case, this still shows how the state lacked information about a minority population and had low capacity in this diverse area. The center was informed of these problems only though a third party.

If, on the other hand, those involved in these abuses were agents of tax farmers, it still shows how limited the state's capacity was due to ethnolinguistic and religious differences, and it shows how local intermediaries possessed informational advantages that can increase their bargaining power vis-a-vis the state. In this case, one intermediary (the Patriarchate) apparently wanted the welfare of the people and asked the state to prevent these problems. However, if the intermediary here was involved in tax collection itself, it would be unlikely that the center would have heard about these abuses and the intermediary could keep pocketing potential revenues that could have accrued in the state's treasury, without the state even hearing about it.

The fact that where the state lacked sufficient control the state had to rely on local intermediaries in counting, ruling and taxing the populations presented a dilemma. The local intermediaries often had an incentive to undercount the populations (Doumani 1994). Many local intermediaries were tax farmers (Barkey 2008; Pamuk 2014), and many were responsible for administering communal taxes.[18] Undercounting would help them obtain cheaper tax farming contracts and decrease the amount of communal taxation demanded by the state. They could increase their profits by pocketing the difference, or if they eased the tax burdens on the local people their position as local notables could be more secure thanks to their popularity by providing better patronage to their clients (Doumani 1994). In short, diversity increased the reliance on intermediaries, intermediaries had an incentive to undercount the population and undercounting resulted in less tax revenues for the state.

---

[18]In communal taxation *(ancemaatin vergi)*, community leaders who were often local notables were responsible to collect a designated amount of revenue from their communities. They were responsible for the allocation of the tax among their own communities, collecting the allocated amount from each individual and delivering the taxes to the government (Özbek 2015).

In his discussion of an Ottoman attempt to count the population of Nablus in the year 1849, Doumani (1994) summarizes the dilemma of having to rely on local intermediaries. Noting that the advisory council of Nablus and the census bureau that had recently been created were the two most important agencies through which the control of the central government could be consolidated through new reforms, he remarks that to be efficient, these institutions had to be staffed by "local leaders" who had a certain degree of knowledge of and influence in the area. However, when these institutions were staffed by them, "...the very social elements who stood to lose from the extension of central control manned the official bureaucratic posts charged with implementing these reforms" (Doumani 1994, p.6). Lacking sufficient knowledge of diverse areas, the state had to rely on local intermediaries to project its rule over these territories and tax the populations. Yet, the very same reliance on the intermediaries made it difficult for the state to exert a higher degree of control and extract more revenues. Echoing this, Çetinsaya (2006) observes that in Ottoman Iraq appointing administrators against the wishes of the local intermediaries created unrest among the intermediaries and the population. However, when the individuals who were preferred by the intermediaries were appointed, "this often appears to have strengthened the notables' power at the expense of the government, and in the long run gave rise to other problems" (Doumani 1994, p.149). Such reliance on and influence of local notables in administrative posts was commonplace in many other Arab regions of the empire (Çetinsaya 2006, p.151).

Having higher bargaining power, the local intermediaries could prevent the state from appointing salaried officials from the center. According to Doumani, in the nineteenth century the local notables in the Nablus advisory council and the Ottoman state constantly "bargained over the boundaries of political authority... The central government had little choice but to cooperate." (1995, p.241). The reason for the high bargaining power of the local intermediaries was their local knowledge. In the bargaining sessions between the Nablus Advisory Council members and the Ottoman state's agents, "...[the] issues revolved around the struggle over access and control of the rural surplus and its disposition and, consequently, over knowledge about the political economy of Jabal Nablus... the Ottoman government tried to gather information about a range of matters, from

population figures to the bidding procedures for commodities collected as taxes-in-kind" (p. 241). However, the local intermediaries in the Advisory council had the upper hand. "In each bargaining session, their responses to requests and admonitions from the central authorities were designed to facilitate their own objectives..." (p.241).

The lack of cooperation by the local intermediaries in Nablus led to a lack of centrally-appointed agents there. Doumani (1995) mentions that the only official in Nablus who was not a local was the head of customs. The difficulty to appoint agents from the center due to the lack of knowledge and resistance from the local intermediaries is not unique to Nablus. The state often had to hand bureaucratic and administrative posts to the local intermediaries in many localities throughout the empire (Köksal 2002; Çetinsaya 2006).

6.3.3   Diversity Undermining Military Conscription

Mass military conscription and mandatory military service, like taxation was a critical part of the state building processes (Tilly 1992). This is why I examine conscription as an alternative dimension of state capacity, which should be similarly affected by diversity and illegibility. By showing that diversity had similar effects on tax collection and military conscription, I aim to demonstrate the robustness of my argument to employing these different dimensions as the dependent variable.

The literature on Ottoman history and the correspondences in the archives suggest that the Ottoman state also encountered difficulties in conscription where the populations were more diverse. During the Tanzimat era, the Ottomans made conscription mandatory for almost all Muslim citizens, with the exception of the inhabitants of the holy cities of Mecca and Madina and those with specific occupations. For non-Muslims mandatory military service was introduced on paper in 1856 but the government preferred not to conscript non-Muslims until 1909, one year after the Young Turk Revolution. Instead, the non-Muslim citizens of the empire continued to pay a poll tax (Zürcher 1998).[19]

---

[19]Even though mandatory military service for non-Muslims was not in place until 1909, the Ottoman administrators had attempted drafting non-Muslims, especially to the navy starting from the 1830s (Gülsoy 2000).

The Ottoman state had difficulty conscripting the minority populations (Zürcher 1998; Talhamy 2011), and had the rely on Muslim and Turkish-speaking portions of the population in the army. This burden disproportionately falling on the Turkish population was openly acknowledged during a debate among the high-ranking Ottoman government officials, by the famous Ottoman statesman Ahmed Cevdet Paşa (Gülsoy 2000, pp.101-2)

There are many instances where the difficulties in conscripting the minorities are evident. The Muslim Roma were never conscripted (Gülsoy 2000) and the recruitment of non-Muslims in the Balkans and Muslims in the Kurdish and Arab-populated regions of the empire often posed significant challenges (Hacısalihoğlu 2007). Due to the lack of censuses in administrative units such as Hejaz *(Hicaz)*, Yemen, Tripoli of Libya *(Trablusgarp)*, Shkoder *(İşkodra)*, Basra, and Hawran *(Havran)*, these areas were exempt from military conscription, even after military conscription was on paper adopted universally in the empire in 1909 (Gülsoy 2000). Almost no Turkish-speakers lived in these provinces. Low-quality censuses also undermined conscription efforts. In a letter to the provinces of Baghdad *(Bağdat)* and Mosul *(Musul)* in 1888, it is mentioned that due to the lack of complete censuses in these regions the majority of the male population were not registered and it is demanded that censuses be started as soon as possible in order to allow conscription (BOA.DH.MKT.1542/96). The Ottoman state was never able to conduct any complete censuses in these provinces until it lost them more than 25 years later during World War I (Karpat 1985a).

Even where the censuses were relatively more successful, minorities often escaped being counted in order to avoid being drafted, or hide after being drafted but before joining the military (Hacısalihoğlu 2007; BOA.A.MKT.MHM. 349/68). In 1909, the Ottoman administrators estimated that about only one tenth of the military-age non-Muslim males were recorded in the population registries sent to the military administrators by the local population offices in the provinces of Syria *(Suriye)*, Aleppo *(Halep)*, Beirut *(Beyrut)* and Adana (Gülsoy 2000, p.140).

If one reason of this disproportionate burden on the core/dominant group was the state being unable to count the minority populations due to their illegibility and the low administrative capacity of the state in more diverse regions, another reason was the higher bargaining power of the local

intermediaries. Intermediaries often helped the possible conscripts hide during population counts to help them avoid conscription (Gülsoy 2000, p.154). They benefited from undercounting the populations not only due to becoming more popular among the people under their control due to relieving tax burdens and allowing many individuals to avoid conscription, but also by having more peasants continuing to work on the land rather than serving long years in the military, themselves often being large landowners (Doumani 1994).

Especially in areas with tribal structures where the local intermediaries still held significant power such as the Arab, Kurdish and Albanian provinces, these intermediaries were responsible for sending conscripts themselves (Zürcher 1998; Hacısalihoğlu 2007). Under such conditions, the tribal chiefs as intermediaries had little incentive to send their fair share of conscripts. With the state not having accurate numbers on the populations in these provinces, these tribal chiefs could get away with not sending many of the young men who were eligible for military service. What is more, they could bribe the government officials with money or land in exchange for erasing the names of potential conscripts from the lists and producing false death certificates (Talhamy 2011).

In the Levant and Syria, where tribal structures prevailed, the Ottoman state encountered significant difficulties in conscripting minority populations such as the Bedouins, Druze and the Alawites (Gülsoy 2000; Talhamy 2011). These areas were under the control of Egypt's ruler Muhammad Ali Paşa for a period during the 1830s and even though Egyptian authorities also initially faced obstacles in taxing and conscripting these populations, but later had managed to solve these issues to some extent (Talhamy 2011).[20] However, one interesting pattern emerges upon a comparison of the achievements of Muhammad Ali's Egypt and the failures of the Ottomans in conscription and taxation in the region. Maoz writes that, even decades after the Ottomans regained control of the area, "the conscripts they were able to raise,... and the taxes they could levy, represented only a small part of what the energetic Egyptians had been able to accomplish" (1968, p.86). Ruling over

---

[20]Muhammad Ali Paşa (or, known in Turkish as Mehmet Ali Paşa), who was assigned as the governor of Egypt by the Ottomans in early nineteenth century, consolidated his power here over the next few decades and challenged the power of the Ottoman state. After his army defeated the Ottoman army during the 1831-3 Ottoman-Egyptian War, he assumed control of Syria and could tax the populations here and conscript soldiers for his own army. The Ottomans regained control of Syria in 1840.

Figure 6.8: Map Showing Draft Evaders in France, 1819-1826

draft evaders

self-mutilation (loss of fingers) to avoid service

draft evaders and self-mutilation
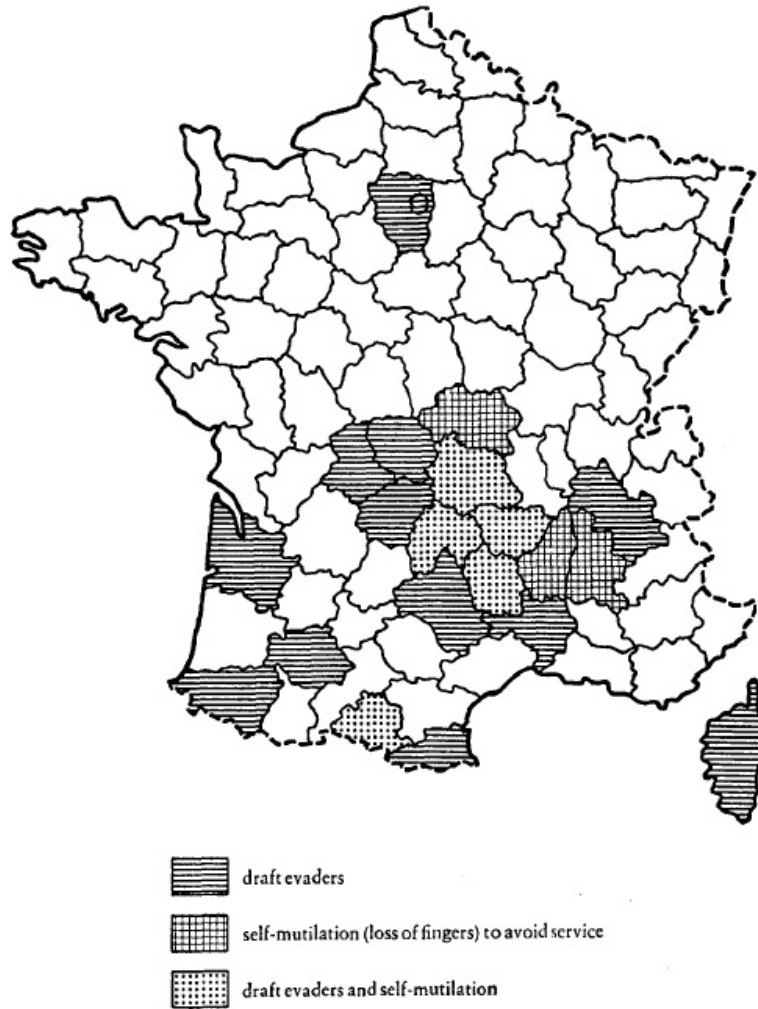
Map 5. Lack of patriotism as reflected in attempts to avoid military service, 1819–1826. SOURCE: E. Le Roy Ladurie, *Anthropologie des conscrits français* (Paris, 1972), pp. 80–81, 138–39.

Note: Taken from Weber (1976, p.106).

these Arabic-speaking populations may have been easier for the Egyptian State, with its capital in Cairo, the population of which was predominantly Arabic-speaking, compared to the Ottoman State, where Arabic was only one of the myriad minority languages.

The two final evidences I will discuss come from nineteenth century France and late-twentieth century Iraq. (Weber 1976; Blaydes 2018). Weber's (1976, p.106) map I present in Figure 6.8 shows instances of draft evasion or self-mutilation in order to escape military service. The former

is probably more relevant to my discussion here as self-mutilation to avoid military service can take place regardless of the levels of legibility. Either way, the patterns in the map are clear. In the South, with a lower proportion of French-speaking population, draft evasion seems to be more common compared to the more-heavily French speaking North of the country (with the exception of Brittany).

Weber interprets these patterns as emanating from different levels of patriotism across French-speaking populations who had embraced the French identity and non-French speaking populations who had not embraced the French identity. However, I argue these patterns are also likely to be a results of the illegibility of the non-French speaking populations to the state's agents. As I have discussed in more detail in Chapter 4, Weber's study on nineteenth century France contains many examples on how the French state's French-speaking agents ran into difficulties in acquiring information about and from the populations that did not speak French (1976).

Similarly, Blaydes (2018, p.279) demonstrates that in the context of late-twentieth century Iraq, draft evasion was more likely among Kurds and the Shi'a Arabs and their evasion was easier compared to Sunni Arabs because the Iraqi state had trouble collecting information about the individuals in these communities and monitoring their activities.

6.4 Conclusion

In this chapter I empirically evaluated Hypotheses 1a and 1b with the local-level fiscal revenue data in the Ottoman Empire between the years 1868 and 1910. The results overall are in line with the expectations in these hypotheses; the wartime changes in fiscal revenues are higher in less diverse administrative units, namely more homogeneous ones and those with higher percentage of Muslim Turkish populations. With this, I provide the main bulk of evidence in this study, that diversity of the population hinders state building.

The finding that the tax revenues in more similar provinces, which had higher proportions of Muslim Turkish speakers, the core/dominant group in this context, increased more than dissimilar provinces indicate that Muslim Turks underwent higher tax burdens of state building. This finding

132

goes against a very intuitive expectation from politics, where one might expect groups in power to tax other groups and benefit themselves. In this case, because taxing members of their own groups is easier and more efficient than taxing the other groups, the core/dominant group taxes its own members.

Using other data sources which include individual-level Property Registers *(Temettüat Defter-leri)* from the 1840s and archival sources, I provided further evidence on how diversity of the population could hinder tax administration and collection, and how Muslims paid more taxes. I also provided evidence from archival sources that indicate how diversity of the population could increase the bargaining power of the intermediaries, and also how diverse populations were more difficult to conscript.

Going forward, the main task to improve the analyses in this chapter can include three items. First, identifying the fiscal revenues for province-years that have missing values, as I show in Table 6.2 can be useful, even though new observations are unlikely to change the overall patterns I identified.

Second, and probably more importantly, identifying more sancak-level fiscal revenue data can improve the sancak-level empirical analyses and help better assess whether the patterns identified with the province-level data also hold at the sancak level.

Finally, collecting further individual-level data from Property Registers and conducting analyses that include data on taxpayers in different geographical areas in the Ottoman Empire of this period can help us understand whether, Muslims on average paid higher taxes compared to non-Muslims in other areas of the empire.

Chapter 7: Diversity Increases the Costs of Investment in State Capacity

In this chapter, I evaluate the two hypotheses on the costs of investment, Hypotheses 2a and 2b. In order to systematically assess these hypotheses I use two sets of data. First, I present evidence from a dataset on the local-level expenditures of the central state in Section 7.1. Next, in Section 7.2, I evaluate evidence from a dataset of Mutasarrıf (administrator of sancak, second-level administrative unit) salaries. Finally, similar to the previous empirical chapters, I present evidence from correspondences in the archives and secondary sources in Section 7.3.

7.1   Evidence from Central State's Local Expenditures

In this section, I present evidence on one aspect of higher costs of investment in fiscal capacity, the local-level expenditures of the state. I first demonstrate that the Ottoman state had to make higher expenditures in provinces with more diverse populations to obtain the same amount of tax revenues. Next, using the same data, I present evidence that in more diverse places the investments focus more on security expenditures that can save the day, while in less diverse places they focus more on items that can increase capacity in the longer-term.

7.1.1   Data and Variables

I use data from local-level expenditures of the Ottoman State, together with the local level fiscal revenue data I introduced in Chapter 6. An ideal test of local-level expenditure patterns during wartime, and check the investment patterns, as establishing better control in an area requires some initial heavy investment (Anscombe 1997, p.44). However, I have not been able to find data on local-level expenditures during wartime. The only data I have been able to locate that covers the central state's expenditures in all provinces was for the fiscal year 1909-10.

As I have argued before, there is likely a 'ratchet effect' where once the state has solved the legibility problems and increased its fiscal capacity in an area, capacity can stay high after the war shock. However, I argue that peacetime investment patterns are still appropriate to test these hypotheses because the state probably needed to make some periodical investment even after the war was over in order to keep capacity from regressing. Without any ongoing investment, the state will not have any updated knowledge about the society and its activities and this can regress its capacity.

The data is from the statistics published by the Ottoman state for the Rumi year 1325 (1909/1910 in Gregorian Years) in the Fiscal Statistics Journal *(Ihsaiyat-i Maliye)*. I exclude the expenses for the ministries of Army and Navy (*Harbiye* and *Bahriye*) because the investments for these ministries and the bureaucrats working under them are arguably less relevant for rendering a population legible and tax collection. However, the results are robust to including them, as I will also present below. The remaining expenses include all the wages paid to the local-level bureaucrats working under various ministries or other departments under the direct control of the government, the local-level expenses of these ministries and departments, and the expenses made for the collection of direct and indirect taxes, such as the wages for those bureaucrats who were responsible for counting and assessing taxes and any other form expenses such procedures cost.

To test Hypotheses 2a and 2b on the costs of investment, I construct a straightforward measure of Expense-to-revenue ratio by dividing the total expenses the state made in a province to the revenues it collected in the same province. This measure indicates how much the state has to invest in a province to extract a unit revenue. The total revenue data are those revenue for the same year and ones that I used in the original dataset, which already comes from the same source, the Fiscal Statistics Journal.

A brief examination of the data reveals that the Hejaz Province (roughly today's Southern Jordan and Western Saudi Arabia) is a very distinct outlier, with an ELF score of 0 but an Expense-to-revenue ratio of 7.37, which means that per each unit tax revenue in this province the state had to make 7.37 units of expenditures, while the mean for the remaining observations is 0.93. This

is a result of this province containing the Muslim holy cities of Mecca and Medina. Its value to the empire was mostly the prestige that it could endow the Ottoman rulers with, thanks to ruling over these holy cities, even though it had little agricultural or commercial activities compared to other Ottoman provinces. The Ottoman state heavily subsidized this province (Ochsenwald 1975). Since this observation is an outlier which is likely to affect the results in a dataset with very small number of observations (N=29 including this observation), I exclude it from the analysis.

I also include several province-level covariates that can confound the relationship between diversity and expense-to-revenue ratio.They include the total population of the province since more populous provinces can be more difficult to administer, a dummy that indicates whether the province has any sea opening, since this can indicate higher trade and export opportunities and therefore more commercialized and monetized agricultural transactions with increased taxation opportunities, a railroad dummy since this railroads make the control of the state easier and tax collection cheaper, the average elevation of the province since it can be more difficult to control territories with high altitudes, a dummy indicating whether the province has any land border to other polities or uncontrolled territory to account for the possibility of increased exports but also that an increased risk of losing a border province may make the state less likely to invest here. Due to the high collinearity caused by other three geographic variables, the distance from the capital (Istanbul), average slope and the total area that the province covers, I had to exclude them from the analysis.

7.1.2   Results: Expense-to-Revenue Ratio

I use OLS Regression to test Hypotheses 2a and 2b about the costs of fiscal capacity building. The results in Table 7.1 using OLS models indicate relationships that are in the expected direction. Model 1 tests the hypotheses without any covariates, while Model 2 includes the covariates. The results in Model 1 indicate that one standard deviation increase in the ELF score increase the Expense-to-revenue ratio by 0.076 (or 7.6 percentage points) and one standard deviation increase

136

Table 7.1: Analysis of the Expenditure-to-revenue Ratio by Diversity for the Fiscal Year 1909-10

	Dependent Variable:	
	Expense to Revenue Ratio	
	(1)	(2)
Ethnolinguistic Fractionalization (ELF)	0.076**	0.062*
	(0.032)	(0.032)
Percent Turkish	−0.070**	−0.106***
	(0.031)	(0.036)
Population (1000)		−0.0001
		(0.0001)
Sea Opening Dummy		0.128*
		(0.071)
Railroad Dummy		−0.110
		(0.066)
Average Elevation (km)		0.163*
		(0.080)
Land Border Dummy		−0.103
		(0.067)
Constant	0.420***	0.328***
	(0.030)	(0.108)
Observations	28	28
R^2	0.320	0.545
Adjusted R^2	0.266	0.386

Note: OLS Regression. Standard errors in parantheses.   *$p<0.1$; **$p<0.05$; ***$p<0.01$

in the Percent Turkish variable decreases the Expense-to-revenue ratio by 7 percentage points. Both are estimated to be significantly different than 0.

In Model 2, the estimate for the Percent Turkish variable in this model is much larger and significantly estimated to be 10.6 percentage points. However, the size of the estimate for the ELF variable declines to 6.2 percentage points and now it misses the conventional levels of significance by a small margin and is estimated to be significant only at the 0.1 level. This estimate is substantially meaningful, the mean of the Expense-to-revenue ratio variable being 0.93 and its standard deviation being 0.52. Unfortunately, the model lacks power with only 28 valid observations and 7 independent variables, making it more difficult to have precise estimates.

Table 7.2: Expense-to-revenue Ratio for All Expense Items for the Fiscal Year 1909-10

	Dependent Variable:	
	Expense to Revenue Ratio	
	(1)	(2)
Ethnolinguistic Fractionalization (ELF)	0.111	0.159**
	(0.081)	(0.073)
Percent Turkish	−0.312***	−0.356***
	(0.079)	(0.081)
Population (1000)		0.001**
		(0.0002)
Sea Opening Dummy		0.253
		(0.163)
Railroad Dummy		−0.118
		(0.151)
Average Elevation (km)		0.268
		(0.183)
Land Border Dummy		0.130
		(0.152)
Constant	0.944***	0.220
	(0.077)	(0.248)
Observations	28	28
R^2	0.427	0.692
Adjusted R^2	0.382	0.584

Note: OLS Regression. Standard Errors in Parantheses. *p<0.1; **p<0.05; ***p<0.01

Expense-to-Revenue Ratio with all Expense Items

For the analyses I report in Table 7.1 I had excluded the expenses for the ministries of the army and the navy. In Table 7.2 I report the results of the analysis for the Expense-to-revenue ratio which includes the expenses for these items, meaning it is constructed using the total of all the expenses. The results in Table 7.2 reveal similar patterns to the results in the original analysis in Table 7.1. The results are still consistent with the expectation that dissimilarity and heterogeneity should increase the expense-to-revenue ratio.

Overall, the results are still consistent with Hypotheses 2a and 2b. All coefficients are in the expected direction. The only coefficient that is not estimated to be statistically significant is the

coefficient for ELF in Model 1, which is in fact estimated to be larger than the coefficient for the same variable in Model 1 of Table 7.1, but is less precisely estimated. This coefficient of 0.111 is substantially still significant, indicating that one standard deviation increase in the ELF measure is estimated to increase Expense-to-revenue ratio by 11.1 percentage points. The lack of precision here, I argue is again likely to be a consequence of lack of power due to small number of observations.

To sum up the discussion so far, I found that the Expense-to-ratio revenue is higher as a province is more heterogeneous and more dissimilar. In other words, the state needs to invest more in more diverse provinces to be able to extract the same unit revenue.

7.1.3 Results: Type of Expenditures by Diversity

One implication of the mechanisms I offer in my argument can relate to the types of expenditures by the state. Where the costs of investment in fiscal capacity is lower, the state should be more likely to focus on items that are more likely to sustain the state's control and increase fiscal capacity in the longer term. In more diverse places, the state's expenditures should be more likely to be save the day and sustain control for the shorter term while not having much benefit in the longer term.

Using the expenditure data for the 1909/10 fiscal year, I identified two spending items that can help the state to increase its control and capacity over the long run, the expenditures on the Ministries of Education (*Maarif*) and Posts and Telegraphs (*Posta ve Telgraf*). These investments are more likely to reflect the peacetime cost-benefit calculations of the decision-makers. Any investment in such infrastructure would be less likely to pay off during the spell of a war and the urgency of raising revenues during wartime, but they are more appropriate to test patterns of costs of investment during peacetime.

Investments in education can increase state capacity in the long run by providing more skilled workforce. In a diverse context education is also very important for capacity building because it allows the state to invest in the education of individuals from minority ethnic groups who speak local languages (BOA.BEO. 4185/313863). Investments in the Ministry of Posts and Telegraphs

also critical because such infrastructural investments indicate a stronger state presence (Acemoglu, Moscona, and Robinson 2016), and increase tax revenues by reducing the costs of communication (Brewer 1989). The variable for measuring long term investments is the proportion of the total of these two expenditures items to all the expenditures that the central state made in a given province.

With the same data, I also calculate the proportion of the total security-related expenditures to the total expenditures. The security forces were often used to help tax collection in the Ottoman Empire in addition to providing security (Anderson 1986; Özbek 2015). Such security expenditures can only help the state control the territory and the populations and enhance tax collection for a very short period of time, such as the security forces being involved in tax collection at that moment and it should not have much effect after the security forces leave.

My expectations, if the mechanisms I offer in my arguments are true, is that under higher similarity and homogeneity the proportion of the long-term investments should be higher, and the proportion of the security investments should be lower.

The results in Table 7.3 are overall consistent with these expectations; all the estimates are in the expected direction although not all are statistically significantly different than 0. In the first two models I test how the two diversity variables predict the proportion of the investments in public goods that can enhance the state's capacity in the longer term. Model 1 is without any province-level covariates. The estimate for the ELF variable is -0.009, indicating that one standard deviation increase in this variable decreases the proportion of the long-term investments by 0.9 percentage points; yet this is significant only at the 0.1 level. When the covariates are included in Model 2, the estimate for the ELF variable increases to -0.014 and is estimated to be significant at the 0.05 level of significance.

In Model 1, without the covariates, the Percent Turkish variable is significantly estimated to be 0.012, meaning that one standard deviation increase in the Percentage of Turks increases the proportion of long-term investments by 1.2 percentage points. However, in Model 2 the size of the estimate declines to 0.01 and is not significantly estimated anymore, missing the significance level of 0.1 by a slim margin.

Table 7.3: Analysis of the Types of Expenditures by Diversity for the Fiscal Year 1909-10

	Dependent Variable:			
	% Longer-Term Investment		% Security Expenditures	
	(1)	(2)	(3)	(4)
Ethnolinguistic Fractionalization (ELF)	−0.009*	−0.014**	0.030	0.039
	(0.005)	(0.006)	(0.026)	(0.024)
Percent Turkish	0.012**	0.010	−0.117***	−0.094***
	(0.005)	(0.006)	(0.026)	(0.025)
Population (1000)		−0.00002		0.0002*
		(0.00002)		(0.0001)
Sea Opening Dummy		0.007		−0.094*
		(0.013)		(0.051)
Railroad Dummy		0.007		0.011
		(0.012)		(0.048)
Average Elevation (km)		0.018		−0.016
		(0.015)		(0.060)
Land Border Dummy		0.003		0.143***
		(0.012)		(0.048)
Constant	0.057***	0.046**	0.627***	0.551***
	(0.005)	(0.020)	(0.025)	(0.079)
Observations	28	28	28	28
R^2	0.299	0.376	0.483	0.747
Adjusted R^2	0.243	0.158	0.441	0.658

*Note: OLS regression. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.*

These estimates are substantively meaningful, even when they are not statistically significant. The mean of the dependent variable is 0.051 and its standard deviation is 0.039. Hence, even when the estimate for the Percent Turkish variable is not statistically significant in Model 2 for example, one standard deviation increase in it predicts 0.26 standard deviations increase in the dependent variable.

Models 3 and 4 test how much the two diversity variables predict the percentage of the security expenditures. Model 3 includes no covariates and Model 4 includes all covariates. The estimate for the ELF variable is 0.03 in Model 3 and 0.039 in Model 4. Neither of these are estimated to be significant at the conventional levels. Still, their sizes are not small. The standard deviation of the dependent variable being 0.192, one standard deviation increase in ELF predicts 0.16 and

0.2 standard deviations increases in Models 3 and 4, respectively. The size of the estimate for the Percent Turkish variable is -0.117 in Model 3 and -0.094 in Model 4. Each of these are estimated to be statistically significant than 0, indicating that one standard deviation in this variable decreases the percentage of security expenditures by 11.7 and 9.4 percentage points respectively.

Overall, all the estimates are in the expected direction but some are not estimated to be significant. This provides some support–albeit not very strong–for the argument that the state should focus its expenditures on items that can increase its control and revenues in the longer term in less diverse contexts while in more diverse contexts the expenditures should be for short-term solutions to the problems of control and tax extraction. The lack of statistical significance for some estimates in some models is very likely due to the very low power, having only 28 observations. Even though the estimates are not always statistically significant, they are substantively significant.

7.2   Evidence from Bureaucrat Salaries

I have argued that hiring more skilled and more experienced bureaucrats and therefore increasing bureaucratic capacity is necessary to increase fiscal capacity. However, in more diverse areas, the state has more limited options to assign bureaucrats and this is a factor that can decrease capacity. In this section, I demonstrate that such difficulties in hiring bureaucrats had repercussions in increasing the salaries the state needed to pay to the bureaucrats who served in more diverse provinces.

7.2.1   Data and Variables

I use another set of data on bureaucrat salaries in the Ottoman Empire, from a list of the salaries paid to the Mutasarrıfs (governors of sancaks, second level administrative units) in the years 1872-1873, as provided by Bouquet (2007). Unfortunately these data are very limited, it only presents the salaries for a small subset of all the sancaks. The salary (in guruş) is the dependent variable, and the ELF and Percent Turkish scores in the sancak are the independent variables.

Table 7.4: List of Sancaks and Mutasarrıf Salaries in the Year 1872-1873

Sancak	Province	Mutasarrıf Salary (in guruş)
Malatya	Diyarbakır	6,500
İçel	Adana	6,550
Bolu	Kastamonu	6,700
Kütahya 1	Hüdavendigar	6,700
Payas	Adana	6,760
Serres (Serez) 1	Thessaloniki (Selanik)	6,760
Hawran (Havran) 1	Syria (Suriye)	6,760
Tekirdağ 1	Edirne	6,760
Kütahya 2	Hüdavendigar	6,760
Gelibolu 1	Edirne	6,770
Hawran (Havran) 2	Syria (Suriye)	7,000
Amasya	Sivas	7,500
Acre (Akka)	Syria (Suriye)	7,500
Balqa (Belka)	Syria (Suriye)	7,500
Lesbos (Midilli)	Med. Islands (Cezayir-i Bahr-i Sefid)	8,000
Tekirdağ 2	Edirne	8,500
Kars	Erzurum	9,000
Debre	Monastir (Manastır) and Prizren	9,000
Canik 1	Trabzon	9,020
Aydın	Aydın	9,020
Hillah (Hille)	Baghdad (Bağdat)	9,020
Sofia (Sofya)	Danube (Tuna)	9,500
Gelibolu 2	Edirne	10,000
Van	Erzurum	10,000
Serres (Serez) 2	Thessaloniki (Selanik)	10,000
Kirkuk (Şehrizor)	Baghdad (Bağdat)	11,000
Niš (Niş)	Prizren	11,000
Karesi 1	Hüdavendigar	11,000
Karesi 2	Hüdavendigar	11,000
Mamüretülaziz	Diyarbakır	11,280
Vidin	Danube (Tuna)	11,500
Rethymno (Resmo)	Crete (Girit)	12,180
Urfa	Aleppo (Halep)	12,500
Beirut (Beyrut) 1	Syria (Suriye)	12,500
Beirut (Beyrut) 2	Syria (Suriye)	12,500
Sfakia (İsfakiye)	Crete (Girit)	12,500
Canik 2	Trabzon	12,500
Trikala (Tırhala)	Ioannina (Yanya)	14,300
Herzegovina (Hersek)	Bosnia (Bosna)	15,000
Heraklion (Kandiye)	Crete (Girit)	17,500

Note: Adapted from Bouquet (2007, pp.368-9). The information on the province under whose jurisdiction the sancak is during this period is from Sezen (2017). Note that some sancaks can be under the jurisdiction of different provinces, or can be independent sancaks in other periods.

I report the Mutasarrıf salary in each sancak in Table 7.4. This table does not include three units Bouquet (2007) lists that I excluded. These are Erbaa, which is a kaza (third-level administrative unit), and Çatalca and Cyprus, which are sancaks under special administration *(müstakil sancak)*. The numbers after the sancak's name indicate those sancaks that had two different governors with different salaries, or where the salary of the governor was changed. For these sancaks, I took the average of the two salaries. I also include the geographical and demographic variables I used in Chapter 4, but I had to exclude the variables that caused high collinearity.

One issue with this sancak-level analysis, similar to the sancak-level analyses in the previous chapters, is that ethnolinguistic diversity measures are not available for many observations. This is why the sample size decreases to 20. What further decreases the sample size to 19 when I use additional covariates is that I do not have geographical variables for the Lesbos *(Midilli)* Sancak and it drops from the analysis in the model where I use the covariates.

7.2.2 Results

I use OLS Regression analysis to evaluate Hypotheses 2a and 2b with these data. Model 1 in Table 7.5 does not include any covariates, and Model 2 includes the sancak-level covariates. The results lend some support to Hypotheses 2a and 2b. According to Model 1, The coefficient of ELF is 914, and is estimated to be statistically significant. This means that one standard deviation increase in the ELF variable is estimated to increase the salary by 914 guruşes. The mean of the dependent variable here being 8,636 and its standard deviation being 2018, this is a substantively large coefficient. One standard deviation increase in the ELF variable increases the dependent variable by 0.45 standard deviations. The coefficient of the Percent Turkish variable in this model is -496, and in the expected direction. Even though this coefficient is not statistically significant, it is substantively still somehow large. One standard deviation increase in the Percent Turkish variable is estimated to decrease the salary by 0.25 standard deviations. Unfortunately, due to the very small number of observations, in this case 20, the model lacks power.

144

Table 7.5: Analysis of Mutasarrıf Salaries by Population Diversity

	Dependent variable:			
	Mutasarrıf Salary (in guruş)			
	(1)	(2)	(3)	(4)
Ethnolinguistic Fractionalization (ELF)	914**	1,249*	914**	1,249
	(420)	(585)	(330)	(647)
Percent Turkish	−496	−639	−496*	−639
	(420)	(524)	(214)	(308)
Area (km. square)		75		75
		(85)		(98)
Average Elevation (km)		−254		−254
		(1,228)		(1,000)
Sea Opening Dummy		1,402		1,402
		(1,577)		(749)
Constant	8,635***	7,107***	8,635***	7,107***
	(410)	(2,079)	(324)	(1,120)
Observations	20	19	20	19
Clustered Standard Errors	No	No	Yes	Yes
R^2	0.262	0.330	0.262	0.329
Adjusted R^2	0.175	0.072	0.176	0.072

Notes: OLS Regression. Standard errors in parentheses. Standard errors clustered at the province-level in Models 3 and 4. *p<0.1; **p<0.05; ***p<0.01.

In the second model, where sancak-level covariates are added, each of the explanatory variables have larger coefficients compared to the first model. Even though the coefficient of the ELF variable increases to 1,249, it is less precisely estimated compared to the first model and is only significant at the 0.1 level. The coefficient of the Percent Turkish variable is still not significantly estimated, even though it is estimated to be -639 here, still substantively large. Once again, this is likely to be a result of the very low number of observations and lack of power here, as the number of observations is only 19.

In Models 3 and 4 of Table 7.5, I replicate Models 1 and 2, but cluster the standard errors at the province level, to account for any within-province correlation of the error terms. In Model 3, the model without any covariates, the coefficient of the ELF variable is still distinguishable from 0

at the conventional levels of significance (0.05), and the coefficient of the Percent Turkish variable is distinguishable from zero at the 0.1 level of significance. In Model 4, when the covariates are added, neither coefficient is estimated to be significantly distinguishable from 0.

7.3 Evidence from Archival and Secondary Sources

Diversity of the population can present the states with challenges of higher costs of investment in fiscal capacity in multiple ways. These can include smaller issues such as having to translate certain legal documents to minority languages and their transportation to the necessary areas (BOA.TFR.I.SL. 124/12400; BOA.TFR.I.MN. 113/11258), or more significant issues such as diversity making investments in communication, transportation, and hiring skilled bureaucrats more costly (Anscombe 1997; Özbek 2015).

The limited budget of Ottoman state tied its hands regarding the necessary investments to enhance the state's control and its capacity. A strong state needs a well-educated workforce, and a strong communication and transportation infrastructure which can allow it to learn about and control the population. The issue of limited resources were evident in Ottoman Iraq, an economically underdeveloped region of the empire, where the state had trouble asserting its control. Çetinsaya mentions many proposals and plans for investment in infrastucture projects in Ottoman Iraq in the nineteenth century, including those that aimed to develop railways, communication and irrigation. However, the lack of resources prevented the financing of these investments (2006, pp.147-148).

As I have mentioned in the previous chapters, the security forces in the empire often were involved in tax collection, especially in areas where the state's control was weak. During the Tanzimat Period, for multiple times the Ottoman state tried to reform the tax collection bureaucracy and hire civilian tax collectors everywhere (Özbek 2015). Investing in a civilian tax collection organization would probably be more costly at the beginning compared to relying on readily available security forces; yet, it is likely to have much higher returns in the longer run, as a more specialized and more skilled tax collection bureaucracy will be more efficient. In spite of these advantages that a potential civilian tax collection bureaucracy offered, Özbek (2015, p.214) notes that there was

not enough resources to invest in a civilian tax collection bureaucracy and the local administrators had to continue relying on security forces for tax collection.
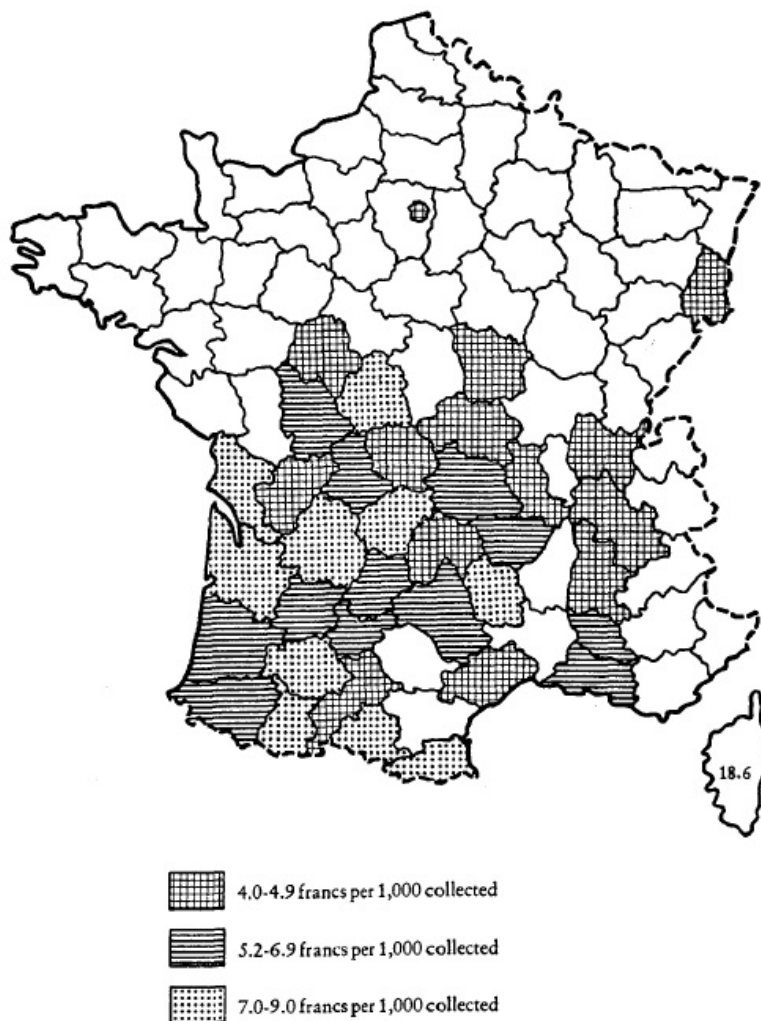
An example of high costs of tax collection under a more diverse population comes from the French context. Weber (1976) provides the map I present in Figure 7.1 to demonstrate that tax collection was more costly in areas with populations that did not speak French and had not adopted the French identity. Clearly, in the Southern half of France in 1834, where the proportion of French speakers were lower compared to the Northern half (again, with the notable exception of Brittany) the costs of tax collection were higher. Similar to the military conscription issue I discussed in Section 6.3.3, Weber interprets this as a result of lack of patriotism in these regions; however, it may well be the population's illegibility due to the population's linguistic diversity.

Weber himself provides an example when in the Haute-Pyrénées region in Southern France "an officer complained that since most of the inhabitants hardly understood any French, one was often forced to have recourse to an interpreter to get any information at all from them" (1976, p.76). Even if the information collection goes perfectly through an interpreter, hiring an interpreter will create higher costs to the state compared to having a population that speaks the language of the majority and being able to send a state agent who does not need an interpreter to obtain information or perform other functions.

One critical area where a state can invest in the longer term to have a more skilled workforce in the longer run is education. From the start of the Tanzimat Era, the Ottoman State invested in many institutions that were designed to train its future bureaucrats and administrators (Findley 1980). The state also invested heavily in primary education during this era (Alkan 2000). However, investment such as primary education is more costly and more difficult under a more diverse population. Different languages can make primary schooling challenging (Weber 1976, pp.81-82), as minority students speaking different languages being schooled in the language of the majority presents problems.

Investing in primary education was a long-term solution. Obtaining benefits from primary schooling probably will start returning yields in years, if not decades. The children who are edu-

Figure 7.1: Map Showing Costs of Collecting Taxes in France in 1834

18.6

4.0-4.9 francs per 1,000 collected

5.2-6.9 francs per 1,000 collected

7.0-9.0 francs per 1,000 collected

Map 6. Lack of patriotism as reflected in the cost of collecting taxes, 1834. SOURCE: Adolphe d'Angeville, *Essai sur la statistique de la population française* (Paris, 1969), p. 342.

Note: Taken from Weber (1976, p.106).

cated need to grow up, join the workforce, and can take positions in the bureaucracy many years after they start their education. Such investments are costly but in the long term necessary. As I argued in Chapter 5 bureaucrats who had knowledge of local languages were critical in increasing the state's bureaucratic and fiscal capacity. The Ottoman administrators were aware of this fact and tried to act accordingly. The reformist governor of Baghdad Midhat Paşa, for example, had issued a decree that each kaza in Baghdad should have a secretary who was schooled in Arabic,

yet whose native language was Turkish (Anscombe 1997, p.50). We cannot learn whether Midhat Paşa achieved to have one such bureaucrat in each kaza. Yet, even if he could, it probably was more costly to hire these people. Having a large pool of potential employees who were schooled in Arabic was probably even more costly as it required long-term investments in education.

Diversity also increased the costs of conducting censuses. From two different documents I have been able to locate, we can observe how censuses could become more costly. In one of these from the province of Van about a census the state was trying to conduct in the year 1894 (BOA.ŞD. 1881/19), it is reported that the census was not completed in Hakkari sancak during the three months when it was supposed to be completed. Hakkari was a very diverse sancak with a very low proportion of Turkish population and above-average heterogeneity levels (0.41). The correspondences in these documents reveal that an extra three months to complete the censuses here was requested by the governor of the province from the ministry of the interior, and the request was accepted by the ministry, as many other areas with similar peculiarities *(hususiyet)* were granted an extra three months. During this extra period, the wages of all those agents who were responsible for the count were going to be paid. In another document, a similar delay in the completion of a census is reported in Edessa *(Vodina)*, in today's Greece (TFR.I.SL. 84/8314).

The lack of information about certain areas created another line of problems for Ottoman administrators due to uncertainties over how much investment would be necessary and how much benefit such investment could return. Regarding the process of the Ottoman state trying to absorb the region of Hasa (roughly today's Eastern Saudi Arabia), Anscombe writes that "Temporarily heavy spending on military and civilian officials could not be avoided in the early stages of any territory's absorption into the empire, and Hasa would certainly prove no exception, due to its unsettled nature and the general Ottoman ignorance about what could be expected there" (1997, p.44).

7.4 Conclusion

In this chapter, I have evaluated Hypotheses 2a and 2b that expected diversity to increase the costs of investment in fiscal capacity. Using data on local-level expenditures of the Ottoman State I found that the State needed to invest more in moire diverse provinces to be able to extract a unit revenue. Using the same set of data, I found that in more diverse provinces the state's investments were more likely to be short-term oriented investments such as security, while in less diverse provinces they were more likely to be long-term oriented investments such as in education or infrastructure. Using another set of data on Mutasarrıf salaries from the years 1872/1873, I found that in more diverse sancaks the state had to pay higher salaries to the Mutasarrıfs, the administrators of these units.

Each of these two different sets of data have significant weaknesses and future work will focus on improving these sets of data. For the expenditures, in ongoing work I am in the process of cleaning data of local-level state expenditures from Provincial Yearbooks. Another task would be to construct a more comprehensive dataset of bureaucratic salaries. The year 1872-1873 would be a good start for this dataset, as I already have many observations and consulting the comprehensive records of bureaucrat biographies of the Ottoman Empire, Sicil-i Ahval Defterleri, I will collect data on the salaries that the Mutasarrıfs in each sancak were paid. This, however, initially requires that I compile a list of all the Mutasarrıfs who served during these years.

Chapter 8: Conclusion

After World War I, the three big empires that until recently had ruled most of Central-Eastern Europe, the Middle East and North Africa, and consisting of many different ethnolinguistic and religious groups splintered into many smaller states. Many of these new states were nation states, they were ethnolinguistically and religiously much more homogeneous. These three empires, Habsburg, Ottoman, and Russian had long been behind their more homogeneous European counterparts such as England, France, and the Netherlands in terms of revenue raising abilities (Karaman and Pamuk 2010).

In this study I focused on one of these empires, the Ottoman Empire, and aimed to demonstrate that its inability in revenue raising and relatedly its inability in successful war making against its European rivals could be explained by the difficulties it faced in obtaining information about its diverse population.

8.1 Findings, Contributions to the Literature, and the General Patterns

The theory and the findings in this study speak to the comparative politics and political economy literatures, and contribute to the topics of the politics of taxation, state building and identity politics. The main findings overall demonstrate that diversity of the population in the Ottoman Empire contributed to its illegibility to the state and the state's agents, which made investment in fiscal capacity costlier in more diverse areas. The higher costs of investment in diverse areas discouraged investment in fiscal capacity. When the Ottoman State needed to extract revenues from the population, it focused its investments in less diverse areas. This resulted in patterns of fiscal capacity building where the wartime increases in the Ottoman state's fiscal revenues were lower in

151

more diverse areas, and where its core/dominant populations, the Sunni Muslim Turks, underwent heavier fiscal burdens.

In order to demonstrate that the mechanism that drives the association between diversity and fiscal capacity building was through diversity making investment in fiscal capacity more costly, I demonstrated that the ratio of the state's expenses to its revenues were higher in more diverse areas. To provide empirical support in favor of other mechanisms that I suggested in the theory, I presented evidence that the Ottoman state was less successful in conducting censuses in more diverse areas, and the diversity of the population disrupts bureaucratic assignment patterns, potentially curtailing bureaucratic capacity. Arguing that strong fiscal capacity is dependent on both high legibility of the population, and high bureaucratic capacity, I interpreted these findings as evidence in favor of the mechanisms I suggest in my argument.

With the results that demonstrate diversity hinders fiscal capacity building and because it increases the costs of investment in fiscal capacity, I contribute to the literature that aims to explain why we observe variation in the state's capacity. These results are in line with Blaydes' (2018) findings where she demonstrates that the Iraqi government's control was weaker in areas with ethnic and religious minorities, Kurds and Shi'a Muslims, because it was more costly to obtain information about these groups and monitor their activities. Similarly, Charnysh (n.d.2022) finds that the higher costs of governing minority groups decreases the state's investment in these groups and the state can extract lower taxes from them. Finally, Gennaioli and Voth (2015) argue that investment in fiscal capacity is more expensive when the populations are more heterogeneous, and they provide cross-country evidence that homogeneity is beneficial for increasing fiscal capacity.

This finding also has implications for the literature on the link between economic development and diversity. High state capacity contributes to economic development (Besley and Persson 2010; Dincecco 2017). Constraining the state's capacity building due to rendering the populations legible, diversity of the population can also constrain economic development. Where the state is unlikely to reap high tax revenues as a result of their investments, it may be discouraged from providing public goods (Charnysh 2022). Lack of investment in public goods can in turn constrain

economic growth. In this sense, illegibility can provide an alternative mechanism why diversity is found to be correlated with worse economic outcomes (Easterly and Levine 1997; Alesina and La Ferrara 2005).

The second main contribution of this study was to identify a distributional consequence of the variation in the costs of investment in fiscal capacity that emanates from the problems of illegibility. The core/dominant group, Sunni Muslim Turks who were in power in the Ottoman Empire undergoing higher tax burdens points to a paradox of maintaining an empire. In a way, being in a politically dominant position imposed costs on this group, where their dominance backfired. Even though this finding may sound counter-intuitive at first glance, it has many analogues in different contexts. It is a similar finding to Kasara's (2007) in which African farmers who had coethnics in power were taxed more heavily because rulers had better control through intermediaries in their home areas, where their coethnics lived. Another example is from Charnysh (n.d.2022), who finds that Muslims paid lower taxes compared to Christians in the Russian Empire, as I discussed above. Other two examples come from other time periods of the Ottoman Empire. Kuran and Rubin (2018), and Cansunar and Kuran (2019) find that the groups in whose favor the law was biased in the Ottoman Empire ended up indirectly being hurt from this bias, for example by having to pay higher interest rates in the market or in terms of weaker property rights. All these examples are instances where being from the identity group in power, which should in theory be advantageous to the members of this identity group, end up hurting them in indirect ways.

A third contribution of the study relates to a very much debated theme in state building, whether wars contribute to stronger states. While there is an extensive literature that argues wars lead to stronger states and increase the states' ability to tax (Hintze 1975 [1906]; Tilly 1992; Besley and Persson 2010; Scheve and Stasavage 2012; Blaydes and Paik 2016), this argument has been found not to be working in many contexts outside of the canonical case of Early Modern Western Europe (Herbst 2000; Heydemann 2000; Centeno 2002). Tilly's account is often oversimplified by assuming that he expects war to directly contribute to state capacity. In fact, Tilly himself expects

wars to lead to stronger states under a certain combination of urbanization and commercialization levels, in a way specifying scope conditions under which wars can contribute to stronger states.

Other works in the literature have identified many other scope conditions for war to result in stronger states. These conditions include the level of urbanization and regime type (Karaman and Pamuk 2013); population density and cooperativeness of the state system (Herbst 2000); nature of political boundaries (Herbst 2000; Heydemann 2000); alternatives to taxation such as easier access to international capital markets (Centeno 2002; Queralt 2019), or printing money (Centeno 2002); limited nature of wars, high levels of interstate conflict, and low levels of initial administrative capacity (Centeno 2002); global strategic considerations, norms of sovereignty, and domestic rents (Heydemann 2000). Identifying that the wartime increases in fiscal capacity of the state are conditional on sufficient ethnolinguistic homogeneity and similarity of the population, this study joins several others that have identified the 'war makes states' (Tilly 1992) relationship to work under certain scope conditions.

Finally, this study offers a contribution to a developing and expanding literature on the origins of the population's legibility or the state's informational capacity (Lee and Zhang 2017; Brambor et al. 2020), by showing that diversity of the population, especially the linguistic differences, make information acquisition more costly and the state and its agents know less about diverse populations. This finding is consistent with other works on legibility from similar as well as different contexts (Scott 1998; Blaydes 2018; Charnysh 2022).

Following Johnson and Koyama (2017) we can consider a few other countries to assess how diversity might have hindered state capacity building efforts. The fact that more homogeneous countries like England and the Netherlands were very successful in earlier periods in increasing their fiscal capacity while more heterogeneous countries like France, Habsburg Empire, Ottoman Empire, Russia and Spain lagged behind these two countries (Karaman and Pamuk 2010; 2013) can be worth thinking about in this sense. While a given war was more likely to have inclined the former, homogeneous group of countries to invest in fiscal capacity earlier, for the latter, more heterogeneous ones this investment was probably more difficult as it was more expensive. This can be

one reason why the investment and the ensuing increases in fiscal capacity may have happened later in the latter group, when technology was more developed (for instance with better transportation and communication networks) and also when they started homogenizing their populations, both of which decrease the costs of investment to exert control over their wider territory and increase fiscal capacity.

Homogenization of populations and building nations constitute a significant step in modernization and state building (Tilly 1975, 1992). An important aspect of these processes of modernization and state building was a unified language (Weber 1976; Anderson 1983; Gellner 1983). Taking a step back and trying to grasp a bigger picture of the process of state building, the findings in this paper are consistent with the view that the top-down nation building and ethnic homogenization projects during times of modernization and state building were probably followed since diverse populations hampered state building, prevented states from achieving higher control of the population and higher tax revenues. It also supports the view that the ethnic, linguistic and cultural homogenization policies that the elite in the Ottoman Empire and the Turkish Republic pursued from the 1910s on were a response to the difficulties in developing state capacity under a diverse society. By the 1910s, the empire was still home to a highly diverse population. So was the early Turkish Republic, even though to a much lesser extent.[1] The transition from the very diverse Ottoman Empire to the much less diverse Turkish republic included policies of assimilation, and ethnic and religious homogenization (Üngör 2011). They involved forcing non-Turkish speakers to start speaking Turkish (Bayar 2011), the Turkification of the economy by transferring wealth from minorities to Turks (Aktar 2001) and a population exchange between Greece and Turkey (Shields 2013). This episode of homogenization of the population in the Ottoman Empire and Turkey also includes acts of violent ethnic cleansing such as the Armenian Genocide (Suny 2015). Acts of violent homogenization of populations were a common theme in the emergence and development of the new nation states in Europe and in nineteenth- and twentieth-century Europe (Ther 2014), and also elsewhere.

---

[1]See the 1927 Census of Turkey.

It is possible to interpret the phenomenon of states homogenizing their populations during state building within Tilly's framework where empires and city states converged towards a new form, the national state (1992).[2] Empires, by definition have diverse populations. On the other hand the end product, the national state, is much more homogeneous. The trajectory from empire to national state Tilly describes, then, may involve an endogenous process where higher homogeneity enabled state development (during mobilization for war), and to achieve state development states may have resorted to homogenizing their populations, often by assimilation and committing atrocities.

8.2 Normative and Policy Implications

Given the previous discussion on how the states struggled to rule and tax diverse populations, and how they may have resorted to homogenization–forcefully and very often violently–in order to rule and tax the populations more efficiently, the implications of the findings in this study may sound bleak, and raise normative concerns. Is it necessary to have non-diverse populations with no cultural, linguistic, and religious differences to achieve stronger states that are more efficient in raising taxes, and providing better public goods?

The answer to this question is no, because of two reasons. First, raising taxes does not always mean those revenues will be used for public good provision, where each individual and each group in the society will equally benefit. The assumption that higher taxes would return as public goods to the taxpayers would be a strong one. The benefits can as well go to the ruling elites, while certain groups are deprived of even the most basic necessities. Stronger states do not always turn out to be beneficial for all groups in the society, and can even deteriorate the welfare conditions of societies in general, or some already worse-off groups (Scott 1976, 1998).

Second, diversity can have other economic benefits, as identified by many different studies. Existence of different ethnic groups can increase the volume of trade and economic growth due to interethnic trade if the different groups live in the same geographical area (Montalvo and Reynal-Querol 2021). It can also boost trade volume through trade with co-ethnics or co-religionists in far-

---

[2]One note of caution is that Tilly does not use this term to refer specifically to 'nation states' or states based on ethnic nationalism.

away places. This was very common in the Ottoman Empire as well as other contexts where some groups' advantages were "complementary and nonreplicable" (Jha 2013, p.807). Many minorities in the Ottoman Empire, such as Armenians, Greeks and Jews more easily traded with their co-ethnics in other regions of the empire or their co-ethnics in other countries thanks to linguistic advantages and networks. Finally, there is also empirical evidence that diversity of populations can be favorable for economic growth thanks to human capital accumulation (Arbatli and Gokmen 2018).

If there is one policy implication we can derive from this study, it is that the potential problems against efficient governance in diverse states can be and should be overcome by more pluralism and democracy, and hence increasing diversity at the ruling elite level. It is more likely that at the root of the problem is having too little diversity among the ruling class, rather than too much diversity among the population.

References

Abramson, Scott F. (2017). "The Economic Origins of the Territorial State". *International Organization* 71.1, pp. 97–130.

Acemoglu, Daron, Jacob Moscona, and James A. Robinson (2016). "State Capacity and American Technology: Evidence from the Nineteenth Century". *American Economic Review* 106.5, pp. 61–67.

Afigbo, A. E. (1972). *The Warrant Chiefs: Indirect Rule in Southeastern Nigeria, 1891-1929*. New York, NY: Humanities Press.

Ahmadi, Shaherzad R. (2021). "Local Ambivalence in the Arabistan-Basra Frontier, 1881-1925". *British Journal of Middle Eastern Studies* 48.3, pp. 436–454.

Akarlı, Engin Deniz (1972). "Ottoman Population in Europe in the 19th Century: Its Territorial, Racial, and Religious Composition". Thesis. University of Wisconsin.

Aktar, Ayhan (2001). *Varlık vergisi ve "Türkleştirme" politikaları*. 5th ed. İletişim Yayınları. Cağaloğlu, İstanbul: İletişim Yayınları.

Akyıldız, Ali (2004). *Osmanlı Bürokrasisi Ve Modernleşme*. Cağaloğlu, İstanbul: İletişim.

Alesina, Alberto, Reza Baqir, and William Easterly (1999). "Public Goods and Ethnic Divisions". *The Quarterly Journal of Economics* 114.4, pp. 1243–1284.

Alesina, Alberto, Arnaud Devleeschauwer, et al. (2003). "Fractionalization". *Journal of Economic Growth* 8.2, pp. 155–194.

Alesina, Alberto and Eliana La Ferrara (2005). "Ethnic Diversity and Economic Performance". *Journal of Economic Literature* 43.3, pp. 762–800.

Alimoski, Safet (2005). "Temettüat Defterlerine Göre Manastir Merkez Kazasinin Sosyo-Ekonomik Durumu". Thesis. Marmara Üniversitesi.

Alkan, Mehmet Ö. (2000). *Tanzimat'tan Cumhuriyet'e Modernleşme Sürecinde Eğitim İstatistikleri 1839-1924*. Ankara: T.C. Başbakanlık Devlet İstatistik Enstitüsü Matbaası.

Anderson, Benedict R. O. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Anderson, Lisa (1986). *The State and Social Transformation in Tunisia and Libya, 1830-1980*. Princeton, N.J.: Princeton University Press.

Andersson, Per F. (2018). "Democracy, Urbanization, and Tax Revenue". *Studies in Comparative International Development* 53.1, pp. 111–150.

Anscombe, Frederick F. (1997). *The Ottoman Gulf: The Creation of Kuwait, Saudi Arabia, and Qatar*. Columbia University Press.

Antonean, Aram (1975). *Balkan Harbi Tarihi*. İstanbul: Sander Yayınları.

Arbatli, Cemal Eren and Gunes Gokmen (2018). *Ethnic Capital and Sub-National Development: Armenian and Greek Legacy in Post-Expulsion Turkey*. SSRN Scholarly Paper ID 3273964. Rochester, NY: Social Science Research Network.

Arpee, Leon (1936). "A Century of Armenian Protestantism". *Church History* 5.2, pp. 150–167.

Arslantaş, Yasin (2017). "Confiscation by the Ruler: A Study of the Ottoman Practice of Müsadere, 1700s-1839". PhD thesis. The London School of Economics and Political Science (LSE).

Arslantaş, Yasin, Antoine Pietri, and Mehrdad Vahabi (2020). "State Predation in Historical Perspective: The Case of Ottoman Müsadere Practice During 1695–1839". *Public Choice* 182.3, pp. 417–442.

Aytekin, E. Attila (2013). "Tax Revolts During the Tanzimat Period (1839–1876) and Before the Young Turk Revolution (1904–1908): Popular Protest and State Formation in the Late Ottoman Empire". *Journal of Policy History* 25.3, pp. 308–333.

Bali, Rifat N. (2005). *The "Varlik Vergisi" Affair*. Istanbul: Isis Press.

Barkey, Karen (1994). *Bandits and Bureaucrats: The Ottoman Route to State Centralization*. Ithaca, N.Y.: Cornell University Press.

— (1996). "In Different Times: Scheduling and Social Control in the Ottoman Empire, 1550 to 1650". *Comparative Studies in Society and History* 38.3, pp. 460–483.

— (2008). *Empire of Difference: The Ottomans in Comparative Perspective*. New York: Cambridge University Press.

Bayar, Yeşim (2011). "The Trajectory of Nation-Building Through Language Policies: The Case of Turkey During the Early Republic (1920–38)". *Nations and Nationalism* 17.1, pp. 108–128.

— (2014). *Formation of the Turkish Nation-State, 1920-1938*. New York, NY: Palgrave Macmillan.

Beramendi, Pablo, Mark Dincecco, and Melissa Rogers (2019). "Intra-Elite Competition and Long-Run Fiscal Development". *The Journal of Politics* 81.1, pp. 49–65.

Bersch, Katherine, Sergio Praça, and Matthew M. Taylor (2017). "Bureaucratic Capacity and Political Autonomy Within National States: Mapping the Archipelago of Excellence in Brazil". In: *States in the Developing World*. Ed. by Miguel Centeno et al. New York, NY: Cambridge University Press, pp. 157–180.

Besley, Timothy and Torsten Persson (2010). "State Capacity, Conflict, and Development". *Econometrica* 78.1, pp. 1–34.

— (2011). *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton, N.J.: Princeton University Press.

Birdal, Murat (2010). *The Political Economy of Ottoman Public Debt: Insolvency and European Financial Control in the Late Nineteenth Century*. London ; New York: Tauris Academic Studies.

Birken, Andreas (1976). *Die Provinzen des Osmanischen Reiches*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Blaydes, Lisa (2018). *State of Repression: Iraq Under Saddam Hussein*. Princeton, N.J.: Princeton University Press.

Blaydes, Lisa and Christopher Paik (2016). "The Impact of Holy Land Crusades on State Formation: War Mobilization, Trade Integration, and Political Development in Medieval Europe". *International Organization* 70.3, pp. 551–586.

Bleaney, Michael and Arcangelo Dimico (2016). "State History, Historical Legitimacy and Modern Ethnic Diversity". *European Journal of Political Economy* 43, pp. 159–170.

Bouquet, Olivier (2007). *Les Pachas Du Sultan: Essai Sur Les Agents Supérieurs De L'état Ottoman (1839-1909)*. Paris: Peeters.

Brambor, Thomas et al. (2020). "The Lay of the Land: Information Capacity and the Modern State". *Comparative Political Studies* 53.2, pp. 175–213.

Brennan, Geoffrey and James M. Buchanan (1980). *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Cambridge, MA: Cambridge University Press.

Brewer, John (1989). *The Sinews of Power: War, Money, and the English State, 1688-1783*. Boston, MA: Unwin Hyman.

Briggs, Asa (1961). "The Welfare State in Historical Perspective". *European Journal of Sociology* 2.2, pp. 221–258.

Çadırcı, Musa (1989). "Türkiye'de Kaza Yönetimi (1840-1876)". *Belleten* 53.206, pp. 237–158.

Cansunar, Asli (2022). "Distributional Consequences of Philanthropic Contributions to Public Goods: Self-Serving Elite in Ottoman Istanbul". *The Journal of Politics* 84.2.

Cansunar, Asli and Timur Kuran (2019). "Economic Harbingers of Political Modernization: Peaceful Explosion of Rights in Ottoman Istanbul". Working Paper. Working Paper.

Centeno, Miguel Angel (2002). *Blood and Debt: War and the Nation-State in Latin America*. University Park, Pa: Pennsylvania State University Press.

Çetinsaya, Gökhan (2006). *Ottoman Administration of Iraq, 1890-1908*. New York, NY: Routledge.

Charnysh, Volha (2022). "Explaining Out-group Bias in Weak States: Religion and Legibility in the 1891/92 Russian Famine". *World Politics* 74.2, pp. 205–248.

Cheibub, Jose Antonio (1998). "Political Regimes and the Extractive Capacity of Governments: Taxation in Democracies and Dictatorship". *World Politics* 50.3, pp. 349–376.

Christensen, Darin and Francisco Garfias (2021). "The Politics of Property Taxation: Fiscal Infrastructure and Electoral Incentives in Brazil". *The Journal of Politics* 83.4, pp. 1399–1416.

Çizakça, Murat (2013). "The Ottoman Government and Economic Life: Taxation, Public Finance and Trade Controls". In: *The Cambridge History of Turkey: The Ottoman Empire as a World Power, 1453-1603*. Ed. by Suraia N. Faroqhi and Kate Fleet. Vol. 2. New York, NY: Cambridge University Press, pp. 241–275.

Clodfelter, Micheal (2008). *Warfare and Armed Conflicts: A Statistical Encyclopedia of Casualty and Other Figures, 1494-2007*. Jefferson, N.C.: McFarland.

Coşgel, Metin M. (2005). "Efficiency and Continuity in Public Finance: The Ottoman System of Taxation". *International Journal of Middle East Studies* 37.4, pp. 567–586.

Creasy, Edward Shepherd (1854). *History of Ottoman Turks; from the Beginning of Their Empire to the Present Time.* Vol. 1. London: R. Bentley.

Croco, Sarah E. and Jessica L. P. Weeks (2016). "War Outcomes and Leader Tenure". *World Politics* 68.4, pp. 577–607.

D'Arcy, Michelle, Marina Nistotskaya, and Ola Olsson (2019). "Mapping the State: Measuring Infrastructural Power through Cadastral Records". Working Paper. Working Paper.

Davison, Roderic H. (1977). "Nationalism as an Ottoman Problem and the Ottoman Response". In: *Nationalism in a Non-national State : The Dissolution of the Ottoman Empire*. Ed. by William Haddad and William Ochsenwald. Columbus, OH: Ohio State University Press.

Deringil, Selim (2003). ""They Live in a State of Nomadism and Savagery": The Late Ottoman Empire and the Post-Colonial Debate". *Comparative Studies in Society and History* 45.2, pp. 311–342.

Dincecco, Mark (2011). *Political Transformations and Public Finances: Europe, 1650-1913*. New York, NY: Cambridge University Press.

Doumani, Beshara B. (1994). "The Political Economy of Population Counts in Ottoman Palestine: Nablus, circa 1850". *International Journal of Middle East Studies* 26.1, pp. 1–17.

— (1995). *Rediscovering Palestine: Merchants and Peasants in Jabal Nablus, 1700–1900*. Berkeley: University of California Press.

Downing, Brian M. (1992). *The Military Revolution and Political Change: Origins of Democracy and Autocracy in Early Modern Europe*. Princeton, N.J.: Princeton University Press.

Doyle, Michael W. (1986). *Empires*. Ithaca, N.Y.: Cornell University Press.

Easterly, William and Ross Levine (1997). "Africa's Growth Tragedy: Policies and Ethnic Divisions". *The Quarterly Journal of Economics* 112.4, pp. 1203–1250.

Ertman, Thomas (1997). *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe*. New York, NY: Cambridge University Press.

Fay, Sidney B. (1950). "Bismarck's Welfate State". *Current History* 18.101, pp. 1–7.

Findley, Carter V. (1980). *Bureaucratic Reform in the Ottoman Empire: The Sublime Porte, 1789-1922*. Princeton, N.J.: Princeton University Press.

Fleming, Katherine Elizabeth (1999). *The Muslim Bonaparte: Diplomacy and Orientalism in Ali Pasha's Greece*. Princeton, N.J: Princeton University Press.

Fukuyama, Francis (2011). *The Origins of Political Order: From Prehuman Times to the French Revolution*. New York, NY: Farrar, Straus and Giroux.

Garfias, Francisco and Emily A. Sellars (2021). "From Conquest to Centralization: Domestic Conflict and the Transition to Direct Rule". *The Journal of Politics* 83.3, pp. 992–1009.

Gat, Azar and Alexander Yakobson (2013). *Nations: The Long History and Deep Roots of Political Ethnicity and Nationalism*. Cambridge, MA: Cambridge University Press.

Gellner, Ernest (1983). *Nations and Nationalism*. New perspectives on the past. Oxford, England: Blackwell.

Gennaioli, Nicola and Hans-Joachim Voth (2015). "State Capacity and Military Conflict". *The Review of Economic Studies* 82.4, pp. 1409–1448.

Göyünç, Nejat (1980). "Trablusgarb'a Ait Bir Layiha". *Osmanlı Araştırmaları/The Journal of Ottoman Studies* 1, pp. 235–256.

Gülsoy, Ufuk (2000). *Osmanlı Gayrimüslimlerinin Askerlik Serüveni*. İstanbul: Simurg.

Güran, Tevfik, ed. (1997). *Osmanlı Devleti'nin Ilk Istatistik Yıllığı 1897*. Ankara: T.C. Başbakanlık Devlet İstatistik Enstitüsü.

— (2014). *19. Yüzyılda Osmanlı Ekonomisi Üzerine Araştırmalar*. Istanbul: Türkiye İş Bankası Kültür Yayınları.

Habyarimana, James et al. (2007). "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101.4, pp. 709–725.

Hacısalihoğlu, Mehmet (2007). "Inclusion and Exclusion: Conscription in the Ottoman Empire". *Journal of Modern European History* 5.2, pp. 264–286.

Haddad, Mahmoud (1994). "The Rise of Arab Nationalism Reconsidered". *International Journal of Middle East Studies* 26.2, pp. 201–222.

Harrison, E. F. (1984). "Report by Mr. E. F. Harrison on the Finances of Turkey". In: *British documents on foreign affairs: reports and papers from the Foreign Office confidential print. Part I, Series B: From the mid-nineteenth century to the First World War. The Near and Middle East, 1856-1914*. Ed. by David Gillard et al. Vol. 7. Frederick, MD: University Publications of America.

Hau, Matthias vom, Jose Peres-Cajias, and Hillel David Soifer (2021). "No taxation without informational foundation: On the role of legibility in tax state development". WIDER Working Paper. WIDER Working Paper.

Herbst, Jeffrey I. (2000). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton, New Jersey: Princeton University Press.

Heydemann, Steven, ed. (2000). *War, Institutions, and Social Change in the Middle East*. Berkeley: University of California Press.

Hintze, Otto (1975). *The Historical Essays of Otto Hintze*. New York, NY: Oxford University Press.

Hirshleifer, Jack (1995). "Anarchy and its Breakdown". *Journal of Political Economy* 103.1, pp. 26–52.

Huber, John D. and Nolan McCarty (2004). "Bureaucratic Capacity, Delegation, and Political Reform". *American Political Science Review* 98.3, pp. 481–494.

Issawi, Charles Philip (1982). "The Transformation of the Economic Position of the Millets in the Nineteenth Century". In: *Christians and Jews in the Ottoman Empire : The Functioning of a Plural Society*. Ed. by Benjamin Braude and Bernard Lewis. Vol. 1. New York, NY: Homes & Meier.

Jha, Saumitra (2013). "Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia". *American Political Science Review* 107.4, pp. 806–832.

Johnson, Noel D. and Mark Koyama (2014). "Tax Farming and the Origins of State Capacity in England and France". *Explorations in Economic History* 51, pp. 1–20.

— (2017). "States and Economic Growth: Capacity and Constraints". *Explorations in Economic History* 64, pp. 1–20.

Karal, Enver Ziya (1943). *Osmanlı İmparatorluğunda İlk Nüfus Sayımı, 1831*. Ankara: Ankara Basım ve Ciltevi.

Karaman, K. Kivanç and Şevket Pamuk (2010). "Ottoman State Finances in European Perspective, 1500–1914". *The Journal of Economic History* 70.3, pp. 593–629.

— (2013). "Different Paths to the Modern State in Europe: The Interaction Between Warfare, Economic Structure, and Political Regime". *American Political Science Review* 107.3, pp. 603–626.

Karaomerlioglu, Asım (2002). "The Peasants in Early Turkish Literature". *East European Quarterly* 36.2, pp. 127–153.

Karaosmanoğlu, Yakup Kadri (2003). *Yaban*. 44-45th ed. İstanbul: İletişim Yayınları.

Karpat, Kemal H. (1978). "Ottoman Population Records and the Census of 1881/82-1893". *International Journal of Middle East Studies* 9.3, pp. 237–274.

— (1982). "Millets and Nationality: The Roots of the Incongruity of Nation and State in the Post-Ottoman Era". In: *Christians and Jews in the Ottoman Empire*. Ed. by Benjamin Braude and Bernard Lewis. New York, NY: Holmes & Meier.

— (1985a). *Ottoman Population, 1830-1914: Demographic and Social Characteristics*. Madison, WI: University of Wisconsin Press.

Karpat, Kemal H. (1985b). "The Ethnicity Problem in a Multi-Ethnic Anational Islamic State: Continuity and Recasting of Ethnic Identity in the Ottoman State." In: *Ethnic Groups and the State*. Ed. by Paul Brass. Totowa, N.J.: Barnes & Noble Books, pp. 94–114.

Kasaba, Reşat (1994). "Social change in the Ottoman Empire". In: *State Power and Social Forces*. Ed. by Joel S. Migdal, Atul Kohli, and Vivienne Shue. New York, NY: Cambridge University Press.

Kasara, Kimuli (2007). "Tax Me If You Can: Ethnic Geography, Democracy, and the Taxation of Agriculture in Africa". *American Political Science Review* 101.1, pp. 159–172.

Kasara, Kimuli and Pavithra Suryanarayan (2015). "When Do the Rich Vote Less Than the Poor and Why? Explaining Turnout Inequality across the World". *American Journal of Political Science* 59.3, pp. 613–627.

Kaya, Alp Yücel and Yücel Terzibaşoğlu (2009). "Tahrir'den Kadastro'ya: 1874 Istanbul Emlak Tahriri ve Vergisi: "kadastro tabir olunur tahrir-i emlak". *Tarih ve Toplum Yeni Yaklaşımlar* 9, pp. 9–58.

Kırmızı, Abdulhamit (2007). *Abdülhamid'in Valileri: Osmanlı Vilayet Idaresi: 1895-1908*. Istanbul: Klasik.

— (2010). "Taming the Governors: The Swinging Pendulum of Power over the Ottoman Provinces in the Nineteenth Century". *The International Journal of Regional and Local Studies* 6.1, pp. 4–23.

Köksal, Yonca (2002). "Local Intermediaries and Ottoman State Centralization: A Comparison of the Tanzimat Reforms in the Provinces of Ankara and Edirne, 1839-1878". PhD thesis. Columbia University.

Kuhnle, Stein and Anne Sander (2010). "The Emergence of the Western Welfare State". In: *The Oxford Handbood of the Welfare State*. Ed. by Francis G. Castles et al. Oxford: Oxford University Press.

Kuran, Timur (2001). "The Provision of Public Goods under Islamic Law: Origins, Impact, and Limitations of the Waqf System". *Law & Society Review* 35.4, pp. 841–898.

— (2004). "The Economic Ascent of the Middle East's Religious Minorities: The Role of Islamic Legal Pluralism". *Journal of Legal Studies* 33.2, pp. 475–515.

— (2011). *The Long Divergence: How Islamic Law Held Back the Middle East*. Princeton, N.J.: Princeton University Press.

Kuran, Timur and Jared Rubin (2018). "The Financial Power of the Powerless: Socio-Economic Status and Interest Rates Under Partial Rule of Law". *The Economic Journal* 128.609, pp. 758–796.

Kushner, David (1977). *The Rise of Turkish Nationalism, 1876-1908*. Totowa, N.J.: Cass.

Kustov, Alexander and Giuliana Pardelli (2018). "Ethnoracial Homogeneity and Public Outcomes: The (Non)effects of Diversity". *American Political Science Review* 112.4, pp. 1096–1103.

La Porta, Rafael et al. (1999). "The Quality of Government". *Journal of Law, Economics, & Organization* 15.1, pp. 222–279.

Lee, Melissa M. and Nan Zhang (2017). "Legibility and the Informational Foundations of State Capacity". *The Journal of Politics* 79.1, pp. 118–132.

Levi, Margaret (1988). *Of Rule and Revenue*. Berkeley, CA: University of California Press.

Lieberman, Evan S. (2002). "Taxation Data as Indicators of State-Society Relations: Possibilities and Pitfalls in Cross-National Research". *Studies in Comparative International Development* 36.4, pp. 89–115.

Livny, Avital (2020). *Trust and the Islamic Advantage: Religious-Based Movements in Turkey and the Muslim World*. Cambridge: Cambridge University Press.

Magiya, Yusuf (2022). "Predatory Rulers, Credible Commitment and Tax Compliance". Working Paper. Working Paper.

Maoz, Moshe (1968). *Ottoman Reform in Syria and Palestine, 1840-1861: The Impact of the Tanzimat on Politics and Society*. Oxford; London: Clarendon Press.

Makdisi, Ussama (2000). *The Culture of Sectarianism: Community, History, and Violence in Nineteenth-Century Ottoman Lebanon*. Berkeley, CA: University of California Press.

— (2002). "Ottoman Orientalism". *The American Historical Review* 107.3, pp. 768–796.

Mann, Michael (1984). "The Autonomous Power of the State: Its Origins, Mechanisms and Results". *European Journal of Sociology* 25.2, pp. 185–213.

Marashlian, Levon (1991). *Politics and demography : Armenians, Turks, and Kurds in the Ottoman Empire*. Cambridge, MA: Zoryan Institute.

Mares, Isabela and Didac Queralt (2015). "The Non-Democratic Origins of Income Taxation". *Comparative Political Studies* 48.14, pp. 1974–2009.

Mares, Isabela and Didac Queralt (2020). "Fiscal Innovation in Nondemocratic Regimes: Elites and the Adoption of the Prussian Income Taxes of the 1890s". *Explorations in Economic History* 77, p. 101340.

Mayshar, Joram, Omer Moav, and Zvika Neeman (2017). "Geography, Transparency, and Institutions". *American Political Science Review* 111.3, pp. 622–636.

McCarthy, Justin and J. Dennis Hyde (1979). "Ottoman Imperial and Provincial Salnames". *Review of Middle East Studies* 13.2, pp. 10–20.

McNamee, Lachlan and Anna Zhang (2019). "Demographic Engineering and International Conflict: Evidence from China and the Former USSR". *International Organization* 73.2, pp. 291–327.

Migdal, Joel S. (1988). *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*. Princeton, N.J: Princeton University Press.

Mitchell, B. R. (2007). *International Historical Statistics: Europe, 1750-2005*. 6th ed. International Historical Statistics. Basingstoke, Hampshire ; New York, N.Y: Palgrave Macmillan.

Montalvo, Jose G. and Marta Reynal-Querol (2005). "Ethnic Diversity and Economic Development". *Journal of Development Economics* 76.2, pp. 293–323.

— (2021). "Ethnic Diversity and Growth: Revisiting the Evidence". *The Review of Economics and Statistics* 103.3, pp. 521–532.

Mou, Hongyu, Licheng Liu, and Yiqing Xu (2022). *panelView: Visualizing Panel Data*. R package version 1.1.8.

Ochsenwald, William L. (1975). "Ottoman Subsidies to the Hijaz, 1877-1886". *International Journal of Middle East Studies* 6.3, pp. 300–307.

Ökte, Faik (1987). *The Tragedy of the Turkish Capital Tax*. Wolfboro, N.H.: Croom Helm.

Ons Volk Ontwaakt (1912). *Ons Volk Ontwaakt: De Godsdiensten op den Balkan*. (Visited on 05/21/2020).

Özbek, Nadir (2010). "Osmanlı İmparatorluğu'nda Gelir Vergisi: 1903-1907 Tarihli Vergi-i Şahsi Uygulaması". *Tarih ve Toplum Yeni Yaklaşımlar* 10, pp. 43–80.

— (2015). *İmparatorluğun Bedeli: Osmanlı'da Vergi, Siyaset Ve Toplumsal Adalet (1839-1908)*. İstanbul: Boğaziçi Üniversitesi Yayınevi.

Özbek, Nadir (2018). "Tax Farming in the Nineteenth-Century Ottoman Empire: Institutional Backwardness or the Emergence of Modern Public Finance?" *The Journal of Interdisciplinary History* 49.2, pp. 219–245.

Özoğlu, Hakan (2001). ""Nationalism" and Kurdish Notables in the Late Ottoman–Early Republican Era". *International Journal of Middle East Studies* 33.3, pp. 383–409.

Pamuk, Şevket (2004). "The Evolution of Financial Institutions in the Ottoman Empire, 1600–1914". *Financial History Review* 11.1, pp. 7–32.

— (2005). *Osmanlı-Türkiye Iktisadi Tarihi 1500-1914*. İstanbul: İletişim.

— (2014). "Fiscal Centralisation and the Rise of the Modern State in the Ottoman Empire". *The Medieval History Journal* 17.1, pp. 1–26.

— (2016). *Türkiye'nin 200 Yıllık Iktisadi Tarihi: Büyüme, Kurumlar ve Bölüşüm*. İstanbul: Türkiye İş Bankası Kültür Yayınları.

— (n.d.). *Geçmiş Yıllara Ait Parasal Büyüklüklerin 1998 Yılı Sonunda Türk Lirası Ve Abd Doları Olarak Eşdeğerleri*.

Pardelli, Giuliana and Alexander Kustov (2022). "When Coethnicity Fails". *World Politics* 74.2, pp. 249–284.

Parker, Geoffrey (1988). *The Military Revolution: Military Innovation and the Rise of the West, 1500-1800*. New York, NY: Cambridge University Press.

Quataert, Donald (1968). "Famine in Turkey, 1873–1875". *Regional Studies* 201, pp. 1–54.

— (2005). *The Ottoman Empire, 1700-1922*. 2nd ed. New York, NY: Cambridge University Press.

Queralt, Didac (2019). "War, International Finance, and Fiscal Capacity in the Long Run". *International Organization* 73.4, pp. 713–753.

Reinkowski, Maurus (2017). *Düzenin Şeyleri, Tanzimat'in Kelimeleri: 19. Yüzyil Osmanlı Reform Politikasının Karşılaştırılmalı Bir Araştırması*. Istanbul: Yapi Kredi Yayinlari.

Roberts, Michael (1956). *The Military Revolution, 1560-1660; an Inaugural Lecture Delivered Before the Queen's University of Belfast*. Belfast: M. Boyd.

Saracoğlu, Rüşdü (2021). *@iriscibre @bestebetul Aynen öyle diyorum. Senelerce Osmanlının kaymağını yiyenler bir gün Cumhuriyette bunun vergisini vermeliler değil mi?* https://twitter.com/RusduSaracoglu/status/1400869940052336644. (Visited on 06/07/2021).

Sarkees, Meredith Reid and Frank Wayman (2010). *Resort to War: 1816 - 2007*. Washington DC: CQ Press.

Scheve, Kenneth and David Stasavage (2012). "Democracy, War, and Wealth: Lessons from Two Centuries of Inheritance Taxation". *American Political Science Review* 106.1, pp. 81–102.

Schumpeter, Joseph A. (1954). "The Crisis of the Tax State". In: *International Economic Papers*. Ed. by Alan Peacock et al. 4. London: Macmillan.

Scott, James C. (1976). *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia*. New Haven, CT: Yale University Press.

— (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.

— (2009). *The Art of not Being Governed: An Anarchist History of Upland Southeast Asia*. New Haven, CT: Yale University Press.

Sezen, Tahir (2017). *Osmanlı Yer Adları*. Ankara: T.C. Başbakanlık Devlet Arşivleri Genel Müdürlüğü.

Sezer Feyzioğlu, Hamiyet (2017). *Bir Osmanlı Valisinin Hazin Sonu: Tepedelenli Ali Paşa İsyanı*. İstanbul: Türkiye İş Bankası Kültür Yayınları.

Shaw, Stanford J. (1978). "The Ottoman Census System and Population, 1831-1914". *International Journal of Middle East Studies* 9.3, pp. 325–338.

Shaw, Stanford J. and Ezel Kural Shaw (1977). *History of the Ottoman Empire and modern Turkey: Volume 2, reform, revolution, and republic: The rise of modern Turkey 1808-1975*. Vol. 11. Cambridge University Press.

Shields, Sarah (2013). "The Greek-Turkish Population Exchange: Internationally Administered Ethnic Cleansing". *Middle East Report* 267, pp. 2–6.

Skaperdas, Stergios (1992). "Cooperation, Conflict, and Power in the Absence of Property Rights". *The American Economic Review* 82.4, pp. 720–739.

Snyder, Richard (2001). "Scaling Down: The Subnational Comparative Method". *Studies in Comparative International Development* 36.1, pp. 93–110.

Soifer, Hillel David (2015). *State Building in Latin America*. New York, NY: Cambridge University Press.

Suny, Ronald Grigor (2015). *"They Can Live in the Desert but Nowhere Else": A History of the Armenian Genocide*. Princeton, N.J.: Princeton University Press.

Suryanarayan, Pavithra (2021). "Status Politics Hollows Out the State: Evidence from Colonial India". Working Paper. Working Paper. Johns Hopkins University.

Talhamy, Yvette (2011). "Conscription among the Nusayris ('Alawis) in the Nineteenth Century". *British Journal of Middle Eastern Studies* 38.1, pp. 23–40.

Tezcür, Güneş Murat, Yusuf Magiya, and Bogdan G. Popescu (2021). "Bureaucratic Shuffling, Regime Type, and Public Good Provision". Working Paper. Working Paper.

Ther, Philipp (2014). *The Dark Side of Nation-States: Ethnic Cleansing in Modern Europe*. New York, NY: Berghahn Books.

Tilly, Charles (1975). "Reflections on the History of European State-Making". In: *The Formation of National States in Western Europe*. Ed. by Charles Tilly. Princeton, N.J: Princeton University Press.

— (1992). *Coercion, Capital, and European States, AD 990-1992*. Cambridge, MA: Blackwell.

Üngör, Ugur Ümit (2011). *The Making of Modern Turkey: Nation and State in Eastern Anatolia, 1913-1950*. Oxford: Oxford University Press.

Weber, Eugen (1976). *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914*. Stanford, CA: Stanford University Press.

White, Eugene N. (2004). "From Privatized to Government-Administered Tax Collection: Tax Farming in Eighteenth-Century France". *The Economic History Review* 57.4, pp. 636–663.

Woods, Henry Charles (1908). *Washed by Four Seas: An English Officer's Travels in the Near East*. London: T. Fisher Unwin.

Yaycıoğlu, Ali (2016). *Partners of the Empire: The Crisis of the Ottoman Order in the Age of Revolutions*. Stanford, CA: Stanford University Press.

Yosmaoglu, Ipek K. (2006). "Counting Bodies, Shaping Souls: The 1903 Census and National Identity in Ottoman Macedonia". *International Journal of Middle East Studies* 38.1, pp. 55–77.

Zamir, Meir (1981). "Population Statistics of the Ottoman Empire in 1914 and 1919". *Middle Eastern Studies* 17.1, pp. 85–106.

Zhang, Nan and Melissa M. Lee (2020). "Literacy and State–Society Interactions in Nineteenth-Century France". *American Journal of Political Science* 64.4, pp. 1001–1016.

Zürcher, Erik Jan (1998). "The Ottoman Conscription System, 1844–1914". *International Review of Social History* 43.3, pp. 437–449.

References for Primary Sources

Başbakanlık Osmanlı Arşivi *(Prime Ministry's Ottoman Archives)*

BOA, A.}MKT, 66/2, H. 22/02/1263 (January 29, 1848)

BOA, A.}MKT.MHM, 349/68, 12/10/1282 (February 28, 1866)

BOA, A.}MKT.UM, 532/59, H. 17/07/1278 (January 18, 1862)

BOA, BEO, 1968/147542, H. 27/09/1320 (December 18, 1902)

BOA, BEO, 4185/31386, H. 18/07/1331 (June 23, 1913)

BOA, DH.MKT, 1373/121, H. 24/01/1304 (October 23, 1886)

BOA, DH.MKT, 1417/5, H. 08/08/1304 (May 2, 1887)

BOA, DH.MKT, 1539/120, H. 29/12/1305 (September 6, 1888)

BOA, DH.MKT, 1542/96, H. 06/01/1306 (September 12, 1888)

BOA, DH.MKT, 1563/25, H.06/03/1306 (November 10, 1888)

BOA, DH.MKT, 1838/16, H. 23/10/1308 (June 1, 1891)

BOA, DH.MKT, 2091/36, H. 26/03/1316 (August 14, 1898)

BOA, DH.MKT, 2485/85, H. 28/01/1319 (May 1, 1901)

BOA, DH.SAİDd, 4/1032

BOA, DH.TMIK.M, 132/42, H. 24/06/1320 (September 28, 1902)

BOA, DH.TMIK.S, 72/30, H. 21/01/1326 (February 24, 1908)

BOA, İ.MVL, 279/10885, H. 12/10/1269 (July 19, 1853)

BOA, İ.MVL, 560/15166, H. 09/05/1283 (September 19, 1866)

BOA, ŞD, 1881/19 H. 27/02/1312 (August 30, 1894)

BOA, TFR.I.A, 36/3546, H. 09/10/1325 (November 15, 1907)

BOA, TFR.I.MN, 113/11258, H. 18/11/1324 (January 3, 1907)

BOA, TFR.I.MN, 138/13707, H. 21/07/1325 (August 30, 1907)

BOA, TFR.I.MN, 148/14704, H. 25/10/1325 (December 1, 1907)

BOA, TFR.I.SL, 84/8314, H. 16/07/1323 (September 16, 1905)

BOA, TFR.I.SL, 124/12400, H. 18/09/1324 (November 5, 1906)

BOA, TFR.KV, 84/8377, H. 07/01/1323 (March 14, 1905)

BOA, Y.EE, 7/12, H. 24/10/1297 (September 17, 1880)

BOA, Y.EE, 58/33, H. 06/04/1327 (April 27, 1909)

BOA, Y.PRK.A, 78/8, H. 05/02/1311 (August 18, 1893)

BOA, Y.PRK.AZJ, 44/68, H. 28/04/1320 (August 4, 1902)

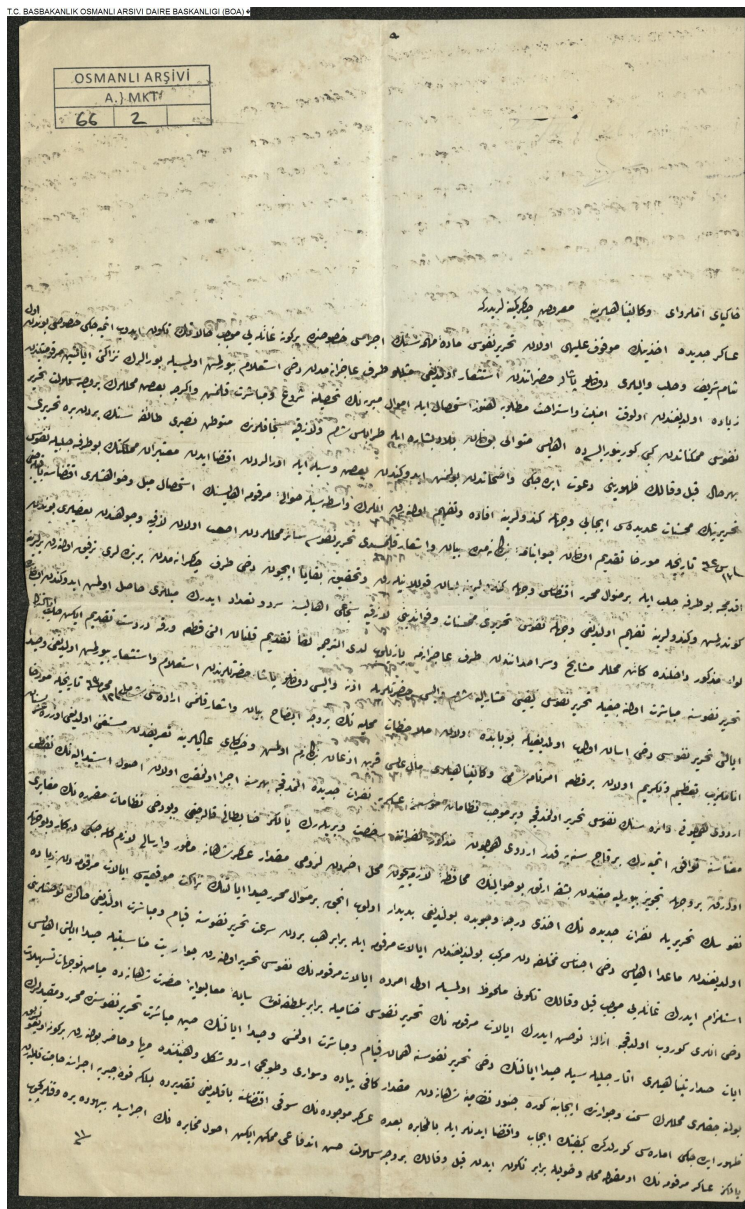**Takvim-i Vekayi** *(Official Gazette of the Ottoman Empire)*

Takvim-i Vekayi, Defa: 238, H. 03/01/1258 (February 14, 1842)

A.1 Letter from Müşir of Sidon *(Sayda)* about the Census (BOA.A.MKT. 66/2)

Figure A.1: BOA.A.MKT. 66/2, page one

A.}MKT.00066.00002.001

A.1.1 Transliteration of BOA.A.MKT. 66/2

Hakipay-i emirlerine ve ali-i vekaletpenahilerine maruz-u çaker-i kemineleridir ki

Asakir-i cedide ahzının mevkuf-u aleyhi olan tahrir-i nüfus madde-i mehamesinin icrası hususunda bir gune gaileyi mucib halatın tekevvün edip etmeyeceği hususu bundan evvel Şam-ı Şerif ve Halep valileri devletlü paşalar hazeratından istişar olunduğu misillü taraf-ı acizanemden dahi istilam buyurulmuş olmasıyla buraların nezaketi eyaletin merfumetinden ziyade olduğundan olvakit emniyet ve istirahat-ı matlube henüz istihsal ile emval-ı miriyenin tahsiline şüru ve mübaşeret kılınmış ve eğer çi bazı mahallerin ber vech-i sühulet tahrir-i nüfusu mümkinattan gibi görünür ise de ahalisi mütevali bulunan Bilad-ı Beşare ile Trablusşam ve Lazkiye sancaklarında mütevattın Nasıri taifesinin birden bire tahriri beher hal kiyl-ü kal zuhurunu davet edeceği vazıhattan bulunmuş idüğünden bazı vesile ile oralardan iktiza eden muteberan-ı memleketin bu tarafa celbiyle nüfus tahririnin muhassenat-ı adidesi icabı vechle kendilerine ifade ve tefhim olunarak anların vasıtasıyla havali-i merkume ahalisinin istihsal meyl ve hahişleri iktizasına bakılacağı 11 Şaban 62 tarihiyle müverrahan takdim olunan cevabname-i bendeganemde beyan ve işar kılınmıştı. Tahrir-i nüfusu sair mahallerden ita olan Lazkiye vücuhundan bazıları bundan akdemce bu tarafa celp ile bir minval-i muharrer iktizası vechile kendilerine lisan kullanılarak ve tahkik-i bakaya için dahi taraf-ı çakeranemden bir bendeleri terfik olunarak yerlerine gönderilmiş ve kendilerine tefhim olunduğu vecihle nüfus tahriri muhassenat ve favaidini Lazkiye sancağı ahalisine serd ve tadad ederek meylleri hasıl olmuş idüğünden ol babda liva-yı mezkur dahilinde kain mahaller meşahih ve seramdanından taraf-ı acizaneme yazılıp ledel-tercüme leffen takdim kılınan altı kıta varaka derdest takdim iken Halep eyaletinin tahrir-i nüfusuna mübaşeret olunacağıyla tahrir-i nüfus liffi müşarileyh Şam valisi hazretleriyle izne ezine valisi devletlü paşa hazretlerinden istilam ve istişar buyurulmuş olduğu ve Sayda eyaleti tahrir-i nüfus dahi asan olup olmadığıyla bu babda olan mülahazat-ı mahalliyenin ber vech-i izah-ı beyan ve işar kılınmış iradesini şamil 12 Muharrem 62 tarihiyle müverrahan enamlerini tazim ve tekrim olan bir kıta emirname-i sami vekaletpenahileri mel-i ali karin-i izan-ı bendeganem olmuş ve hak-i pay-i alilerine tariften müstağni olduğu üzere sernişan ordu-yu hümayun dairesinin nüfusu tahrir olunmadıkça ve ber mucib-i nizamat-ı müessese-i askeriye neferat-ı cedide alınmadıkça beher sene icra olunmakta olan usul-u istibdaliyenin tanzimi meza sene tevakkuf etmeyerek bir kaç seneye kadar ordu-yu hümayun-u mezkur neferatına ruhsat verilerek yalnız zabıtanı kalacağı ve bu dahi nizamat-ı mukarrerenin mugayiri olarak bir vecihle tecviz buyurulmayacağından başka artık bu havalinin muhafaza-yı lazımıyyesiçün mahall-i uhradan lüzumu miktar asakir-i şahane memur ve irsali lazım geleceği derkar ve bu cihetle nüfusun tahririyle nefaret-i cedidenin ahzı derece-i vücubda bulunduğu bedidar olup ancak bir minval-i muharrer Sayda eyaletinin nezaket-i mevkiyesi eyalet-i merkumeden ziyade olduğundan maada ahalisi dahi ecnas-ı muhtelifeden mürekkep bulunduğundan eyalet-i merkume ile beraber hep birden sürat tahrir-i nüfusuna kıyam ve mübaşeret olunduğu halde tevahhuşlarını istilzam ederek gaileyi mucip kil-ü kalin tekevvünü melhuz olmasıyla evvel emirde eyalat-i merkumenin nüfusu tahrir olunarak civariyet münasebetiyle Sayda eyaleti ahalisi dahi anları görüp oldukça izale-i tevehhün ederek eyalat-i merkumenin tahrir-i nüfusu hitamıyla beraber bil-lutf-i teala saye-i mealivayi-i hazret-i şahanede meyamin tevcihat-ı teshilat ayat-ı sadaretpenahileri asar-ı celilesiyle Sayda eyaletinin dahi tahrir-i nüfusuna heman kıyam ve mübaşeret olunması ve Sayda eyaletinin

hın-ı mübaşeret-i tahrir-i nüfusunda muharrer ve mukayyedlerin bulunacakları mahallerin semt ve civarına icabına göre cünud-u nizamiye-i şahaneden miktar-ı kafi piyade ve süvari ve topçu ordu şekil ve heyetinde müheyya ve hazır bulunarak bir gune uygunsuzluk zuhur edeceği emaresi görüldükte keyfiyetin icab ve iktiza edenler ile bilamuhabere badehu asker-i mevcudenin sevki iktizasına bakıldığı takdirde belki kuvve-yi cebriye icrasına hacet kılınarak yalnız asakir-i merkumenin o mukime mahalle ve havliyle beraber tekevvün eden kil-ü kalin ber vech-i suhulet hüsn-ü indifaı mümkün iken usul-u muhaberenin icrasıyla beyhude yere vakitler geçip sonra kil-ü kal başka yerlere dahi sirayetle maslahatın büyümesini mucib olacağı ecilden olvecihle o makule halet vukuunda usul-u muhaberenin icrasından sarf-ı nazarla asakir-i merkume ile memur bulunacak zatın işin gelişine ve maslahatın revişine göre heman tez elden icabının ifası hususuna kıyam ve müsaraat eylemesiçün ordu-yu hümayun-u mezkur müşiri devletlü paşa hazretlerine işar ve irade buyurulmuş ve sonra tekrar bir takım tekalifatın ihtiyarına hacet kalmamak üzere Sayda eyaleti tahrir-i nüfusu sırasında neferat-ı cedidenin dahi ahz kılınması ve sevahilde kain mahallerin hın-ı tahrir-i nüfusunda süfün-ü donanma-yı hümayun o makule mahaller iskele ve civarlarında bulunmasında tesirat başka olacağına mebni iki kıta sefine-i şahanenin irsaline müsade-i aliye erzan buyrulması mütalaa-ı kasire-yi acizaneme münasip gibi görünüyorsa da Sayda eyaleti nüfus tahririnin eyalet-i merkume nüfus tahriri hitama karib olacağı vakte talik olmasını veyahut heman şimdiden kıyam ve mübaşeret kılınması mütevakkıf rey ve irade-yi ali olduğu ve Sayda eyaletinin tahrir-i nüfusuna beden ve mübaşeret olunması emir ve ferman buyurulduğu vakit iktiza edecek talimat-ı seniyenin tesyarına dahi ihsan buyurulması lazıme-yi halden idiği beyanıyla ariza-yi çakeranem terkim ve takdimine ictira kılınmıştır. Mütaallik mümataallüm-ü aliye-yi vekaletpenahileri buyuruldukta ol babda ve her halde emrü ferman hazret-i men lehül emrindir.

22 Safer 64

Sayda müşiri bende

Mehmed Kamil

A.2 Telegram Asking Bureaucrats to be Assigned to Kozani *(Kozana)* to Speak Local Languages (BOA.TFR.I.MN. 138/13707)

Figure A.3: BOA.TFR.I.MN. 138/13707

TFR.I.MN.00138.13707.001

A.2.1 Transliteration of BOA.TFR.I.MN. 138/13707
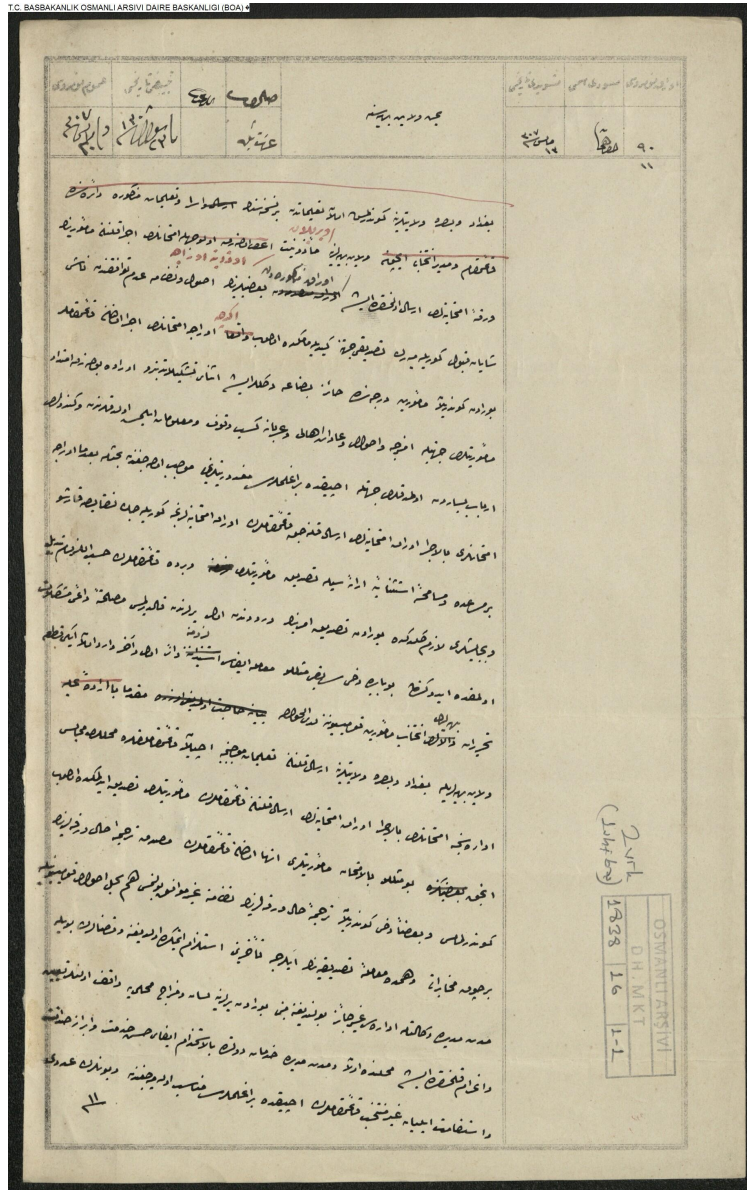
Huzur-u sami-yi müfettiş-i efhamiye

Defterdarlığa intisab ve Kozana'ya izam kılınan haric-i şehir vergi katibinin lisan-ı mahalliye aşina olmadığı anlaşılmış ve bu kazada adedi kesir olan ve sırf Rumca konuşan karyelerin tadilatını yapmakta duçar-ı müşkilat olacağı derkar bulunmuş olunmasına tahsildarların intihabında lisan-ı Rum'dan maada lisan-ı mahalliyeye aşina olanların tercihi hakkındaki müsaade-i nizamiyenin virgü katiplerinin intihabına dahi teşmiliyle gerek Kozana ve gerek diğer emsali kazalarında kain kurada hakkıyla iş görebilecek vergi ketebesi tedariki esbabının istikmali zımnında icab edenlere evamir-i lazime ita buyurulması her ? müsade-i celile-i cenab-ı daver itibaridir denile?
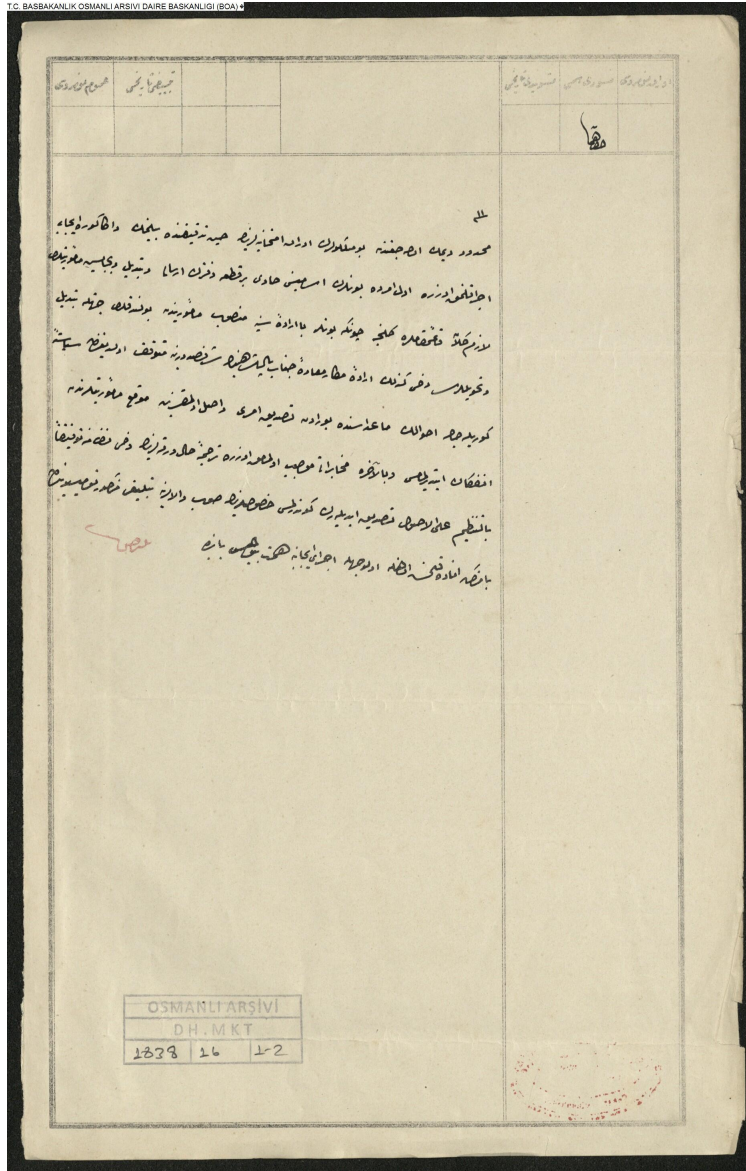
15 Ağustos 323

Cemal

A.3 Telegram to the Province of Yemen About Not Dismissing the Bureucrats Who Failed the Exam if They Speak Arabic or are Familiar with Arab Culture (BOA.DH.MK. 1838/16)

Figure A.4: BOA.DH.MK 1838/16, page one

DH.MKT.01838.00016.001

A.3.1 Transliteration of BOA.DH.MK 1838/16

Umum numerosu	Tebyizi tarihi	?	?	Yemen vilayet-i behiyesine	Müsevvidi tarihi	Müsevvidi ismi	Evrak numerosu
20 Mayıs 307	23 Şevval 1308		İzzet Bey		13 Mayıs 307	?	90 11

Bağdat ve Basra vilayetlerine gönderilmiş olan talimattan bir nüshasının ~~irsal ve~~ isra ve talimat-ı mezkure dairesinde ~~kaimmakam ve mudur intihabi icun~~ vilayet-i behiyelerine (verilen) mezuniyet (üzerine oraca) ~~ita olunarak ol vecihle~~ imtihanlar icra kılınan memurinin varaka-yı imtihaniyelerin irsal olunmakta ise de ~~evrak-i musaddereden~~ (evrak-ı mezkureden) bazılarının usul ve nizama adem-i tevafukundan naşi şayan-ı kabul görülemeyerek tasdiki cihetine gidilememekte olup ~~vakia~~ (eğerçi) oraca imtihanları icra olunan kaimmakamlar buradan gönderilen memurin derecesine haiz bidaa değiller ise de esna-yı teşkilattanberü orada bulunrarak imtidar işdar-ı memuriyetleri cihetiyle eğerçi ve usuller ve adet-i ahali ve urban-ı kesb-i vükuf ve malumat esliha olduklarından ve kendileri erbab-ı lisandan olmadıkları cihetle açıkta bırakılmaları mağduriyetlerini mucib olacağından bahisle badema oraca imtihanları bila-icra evrak-ı imtihaniyeleri irsal kılınacak kaimmakamların evrak-ı imtihaniyelerince görülecek tezayıka karşı bir müsaade ve müsamaha-yı istisnaiye-i ira'esiyle tasdik-i memuriyetleri ve bir de Kaimmakamların hasbel luzum tebdil ve becayişleri lazım geldikte buradan tasdik emrinin vürudundan evvel yerlerinden kaldırılması maslahat-i da'i-i müşkilat olmağda idüğünden bu babda dahi seri-i misillü muamele ifası lüzumuna dair evvel ve ahir varid olan iki kıta tahrirat-i (behiyeleri) ~~delalet~~ intihab-ı memuriyet komisyonuna ledel-hulle ~~beyan-i hacet olmadigi uzere~~ mukaddema ~~bairade-i aliye~~ vilayet-i behiyyeleriyle Bağdat ve Basra vilayetlerine irsal kılınan talimat mucibince açılan kaimmakamlıklara mahalli mecalis idaresince imtihanları bila-icra evrak-ı imtihaniyeleri irsal kılınan kaimmakamların memuriyetleri tasdik edilmekte olup ancak ~~bazilarinda~~ bu misillü bilimtihan memuriyetleri inha olunan kaim-makamların musaddık tercüme-i hal varakalarının gönderilmemesi ve bazan dahi gönderilen tercüme-i hal varakalarının nizamına gayr-ı muvafık bulunması hem sicil-i usul komisyonuyla bir çok muhaberatı ve hem de muamele-i tasdikiyenin aylarca tehirini istilzam etmekte olduğuna ve kazaların böyle müddet-i müdüre vekaletle idaresi gayr-ı caiz bulunduğuna mebni buradan yerlerine lisan ve mizac-ı mahalliye vakıf olanlar tayin ve izam kılınmakta ise de mahalinde olan ve müddet-i müdürde hizmet-i devlette bilistihdam ifası hüsn-ü hizmet ve ibraz-ı sadakat ve istikamet eyleyene gayr-ı müntehib kaimmakamların açıkta bırakılmaları münasip olamayacağından ve bunların adedi mahdud demek olacağından bu misillülerin evrak-ı imtihaniyelerinin hıyn-i tetkikinde beyanımın ve ona göre icabını icra kılınmak üzere evvel emirde bunların esamisini havi bir kıta defterin irsali ve tebdil ve becayiş-i memuriyetlere lazım gelen kaymakamlara gelince çünkü bunlar ba-irade-i seniyye mensup memurinden bulundukları cihetle tebdil ve tahvilleri dahi kezalik irade-yi makarr-ı müsaade-yi cenab-ı şerefsudurdan mütevakkif olduğuna siyaseten görülecek ahvalin maadasında buradan tasdik emri dahil olmaksızın mevkii-yi memuriyetlerinden infikak ettirilmemesi ve bila-hare muhaberatı mucib olmamak üzere tercüme-yi hal varakalarının dahi nizamına tevfikan bil-

tanzim alelusul tasdik edilerek gönderilmesi hususlarının savb-ı valasına tebliği mezkur komisy-onun ba-tezkere ifade kılınmış olmağla olvecihle icra-yı icabına himmet buyurulması babında.[1]

---

[1]The words that are crossed out in the original source are written as crossed out in the transliteration and the words that were added later are written in parentheses.

Appendix B: Occupations Coded as State Officials in Temettü Registers

Alay Emini, Alay Katibi, Asker, Askeri Maaşdan (?), Binbaşı, Dergah-ı Ali Kapucubaşısı, Hacegan-ı Divan-ı Hümayûn, Kaymakam, Kocabaşı, Kol Ağası, Miralay, Müdür, Muhzır, Mülazım, Nüfus Nazırı, Sandık Emini, Sayıcı, Tabur Katibi, Timarlı Zabıta, Timarlı Zabıta Ketebesi, Topçu Alayı Katibi, Yüzbaşı, Zabıta.[1]

---

[1]I write each category verbatim as it is transliterated in the source.

Appendix C: List of Sancaks in the Sancak-Level Revenue Analysis

Table C.1: List of Sancaks for which Data are Available in Section 6.1.5

Sancak	Province	Data Exists (Period)
Adana	Aleppo, later Adana	All
Akka	Syria, later Beirut	All
Aleppo	Aleppo	1877-1896, 1897-1910
Amara	Baghdad, later Basra	1877-1896, 1897-1910
Amasya	Sivas	1868-1876, 1877-1896
Ankara	Ankara	1868-1876, 1877-1896
Aydın	Aydın	1877-1896, 1897-1910
Balqa	Syria, later Beirut	1877-1896, 1897-1910
Beirut	Syria, later Beirut	1877-1896, 1897-1910
Biga/Kale-i Sultaniye	Independent Sancak, Hüdavendigar, Mediterranean Islands, Karesi	All
Benghazi	Tripoli of Libya, later Independent Sancak	1877-1896, 1897-1910
Bolu	Kastamonu	All
Burdur	Konya	All
Canik	Trabzon, at times Independent Sancak	All
Çatalca	Independent Sancak	1877-1896, 1897-1910
Cebel-i Bereket/Payas	Aleppo, later Adana	All
Damascus	Syria	1868-1876, 1877-1896
Denizli	Aydın	1877-1896, 1897-1910

184

Table C.1 – continued from previous page

Sancak	Province	Data Exists (Period)
Dersim	Dersim, later Mamüretülaziz	1877-1896, 1897-1910
Gelibolu	Edirne	All
Gümüşhane	Trabzon	All
Hama	Syria	1868-1876, 1877-1896
Hamit	Konya	All
Harput/Mamuretülaziz	Diyarbakır, later Mamüretülaziz	All
Hawran	Syria	1868-1876, 1877-1896
İçel	Konya, later Adana	All
İzmir	Aydın	1877-1896, 1897-1910
İzmit	Independent Sancak	All
Jerusalem	Syria, later Independent Sancak	All
Karahisar-ı Şarki	Sivas	1868-1876, 1877-1896
Karesi	Hüdavendigar, at times Karesi, and Independent Sancak	1877-1896, 1897-1910
Kastamonu	Kastamonu	All
Kayseri	Ankara	1868-1876, 1877-1896
Kengiri (Çankırı)	Kastamonu	All
Kırkkilise	Edirne	1877-1896, 1897-1910
Kırşehir	Konya, later Ankara	1868-1876, 1877-1896
Konya	Konya	All
Kozan	Aleppo, later Adana	All
Latakia	Syria, later Beirut	1877-1896, 1897-1910
Lazistan	Trabzon	1877-1896, 1897-1910
Malatya	Diyarbakır, later Mamüretülaziz	All
Maraş	Aleppo	1877-1896, 1897-1910

Table C.1 – continued from previous page

Sancak	Province	Data Exists (Period)
Menteşe	Aydın	1877-1896, 1897-1910
Muntafiq	Baghdad, later Basra	1877-1896, 1897-1910
Nigde	Konya	1877-1896, 1897-1910
Saruhan	Aydın	1877-1896, 1897-1910
Sinop	Kastamonu	All
Sivas	Sivas	1868-1876, 1877-1896
Teke	Konya	All
Tekfurdağı	Edirne	All
Tripoli	Syria, later Beirut	1877-1896, 1897-1910
Trabzon	Trabzon	All
Urfa	Aleppo	1877-1896, 1897-1910
Yozgat	Ankara	1868-1876, 1877-1896
Zor	Aleppo, later Indepdendent Sancak	1877-1896, 1897-1910

Appendix D: Robustness Checks

Table D.1: Generalized Difference-in-Differences Analysis of the Wartime Increases in Fiscal Revenues with Alternative Population Data

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	26.910***
	(1.317)
Percent Turkish	6.536***
	(1.264)
Post-1877 Dummy	5.093
	(3.508)
Post-1897 Dummy	6.440*
	(3.339)
ELF * Post-1877 Dummy	−2.140**
	(0.900)
ELF * Post-1897 Dummy	−1.921**
	(0.853)
Percent Turkish * Post-1877 Dummy	4.226***
	(0.884)
Percent Turkish * Post-1897 Dummy	4.842***
	(0.864)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	734
R^2	0.869
Adjusted R^2	0.855

Note: OLS Regression. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table D.2: Generalized Difference-in-Differences Analysis of the Wartime Increases in Fiscal Revenues with Alternative Revenue Data

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	28.918***
	(1.250)
Percent Turkish	7.357***
	(1.200)
Post-1877 Dummy	4.098
	(3.329)
Post-1897 Dummy	6.291**
	(3.130)
ELF * Post-1877 Dummy	−3.059***
	(0.855)
ELF * Post-1897 Dummy	−2.504***
	(0.806)
Percent Turkish * Post-1877 Dummy	3.937***
	(0.839)
Percent Turkish * Post-1897 Dummy	3.906***
	(0.817)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	736
R^2	0.891
Adjusted R^2	0.879

Note: OLS Regression. Standard errors in parentheses.   *$p<0.1$; **$p<0.05$; ***$p<0.01$

Table D.3: Panel Data Analysis of Wartime Changes in Fiscal Revenues

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Post-1877 Dummy	1.792**
	(0.859)
Post-1897 Dummy	5.574***
	(0.796)
ELF * Post-1877 Dummy	−2.681***
	(0.850)
ELF * Post-1897 Dummy	−1.888**
	(0.803)
Percent Turkish * Post-1877 Dummy	2.467***
	(0.835)
Percent Turkish * Post-1897 Dummy	3.312***
	(0.810)
Observations	736
R^2	0.190
Adjusted R^2	0.151

Note: Within-effects Regression. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table D.4: Generalized Difference-in-Differences Analysis of Wartime Changes in Fiscal Revenues for only Provinces that Exist in all Three Periods

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	30.707***
	(1.322)
Percent Turkish	6.068***
	(1.290)
Post-1877 Dummy	6.421*
	(3.410)
Post-1897 Dummy	5.559*
	(3.278)
ELF * Post-1877 Dummy	−2.232**
	(0.897)
ELF * Post-1897 Dummy	−1.864**
	(0.893)
Percent Turkish * Post-1877 Dummy	4.152***
	(0.899)
Percent Turkish * Post-1897 Dummy	4.352***
	(0.900)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	683
R^2	0.881
Adjusted R^2	0.867

Note: OLS Regression. Standard errors in parentheses.   *$p<0.1$; **$p<0.05$; ***$p<0.01$

Table D.5: Generalized Difference-in-Differences Analysis of Wartime Changes in Fiscal Revenues Excluding the Three Middle Anatolian Provinces Affected by the Famine of mid-1870s

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	23.978***
	(1.329)
Percent Turkish	9.487***
	(1.280)
Post-1877 Dummy	5.998*
	(3.564)
Post-1897 Dummy	7.325**
	(3.301)
ELF * Post-1877 Dummy	−3.119***
	(0.944)
ELF * Post-1897 Dummy	−2.574***
	(0.885)
Percent Turkish * Post-1877 Dummy	4.574***
	(1.086)
Percent Turkish * Post-1897 Dummy	4.347***
	(1.032)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	639
R^2	0.908
Adjusted R^2	0.897

Note: OLS Regression. Standard errors in parentheses.   *p<0.1; **p<0.05; ***p<0.01

Table D.6: Generalized Difference-in-Differences Analysis of Wartime Changes in Fiscal Revenues for the Sample where Warzones are Excluded

	Dependent Variable:
	Revenue Per Male Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	12.475***
	(2.608)
Percent Turkish	−14.685**
	(5.977)
Post-1877 Dummy	8.707***
	(3.192)
Post-1897 Dummy	5.838**
	(2.932)
ELF * Post-1877 Dummy	−1.652*
	(0.856)
ELF * Post-1897 Dummy	−2.966***
	(0.783)
Percent Turkish * Post-1877 Dummy	2.262***
	(0.803)
Percent Turkish * Post-1897 Dummy	3.484***
	(0.759)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	644
R^2	0.910
Adjusted R^2	0.899

Note: OLS Regression. Standard errors in parentheses. $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table D.7: Generalized Difference-in-Differences Analysis of Wartime Changes in Fiscal Revenues for the Post-1871 Sample

	Dependent Variable:
	Revenue Per Capita (in 1998 USD)
Ethnolinguistic Fractionalization (ELF)	14.575***
	(3.880)
Percent Turkish	9.903**
	(4.692)
Post-1877 Dummy	6.863**
	(2.836)
Post-1897 Dummy	5.957**
	(2.680)
ELF * Post-1877 Dummy	−1.903**
	(0.930)
ELF * Post-1897 Dummy	−2.279***
	(0.709)
Percent Turkish * Post-1877 Dummy	2.495***
	(0.924)
Percent Turkish * Post-1897 Dummy	3.346***
	(0.719)
Province Fixed Effects	Yes
Year Fixed Effects	Yes
Observations	646
R^2	0.915
Adjusted R^2	0.904

Note: OLS Regression. Standard errors in parentheses. $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$