

DOCTORAL SCHOOL 465- ÉCONOMIE PANTHÉON SORBONNE

PARIS SCHOOL OF ECONOMICS

Ph.D. Thesis Submitted for the Degree of Doctor of Philosophy in Economics

Analysis and Policy in Economics

Lydia Assouad

**Essays on the Political Economy of
Development of the Middle East**

Supervised by Thomas PIKETTY et Ekaterina ZHURAVSKAYA (PSE)

Defense date : June 20, 2022

Referees 1. Daron ACEMOĞLU, MIT
2. Noam YUCHTMAN, LSE

Jury 1. Daron ACEMOĞLU, MIT
2. Thomas PIKETTY PSE
3. Mohamed SALEH TSE
4. Noam YUCHTMAN, LSE
5. Ekaterina ZHURAVSKAYA PSE

ÉCOLE DOCTORALE 465- ÉCONOMIE PANTHÉON SORBONNE

ÉCOLE D'ÉCONOMIE DE PARIS

Thèse pour l'obtention du titre de Docteur en Sciences Économiques

Discipline : Analyse et Politique Économique

Lydia Assouad

**Essais sur l'économie politique du
développement du Moyen-Orient**

Thèse dirigée par : Thomas PIKETTY et Ekaterina ZHURAVSKAYA (PSE)

Date de la soutenance : le 20 juin 2022

Rapporteurs 1. Daron ACEMOĞLU, MIT
2. Noam YUCHTMAN, LSE

Jury 1. Daron ACEMOĞLU, MIT
2. Thomas PIKETTY PSE
3. Mohamed SALEH TSE
4. Noam YUCHTMAN, LSE
5. Ekaterina ZHURAVSKAYA PSE

ACKNOWLEDGMENTS

This thesis would not have been possible without the support of many people I was lucky to meet and learn from. First of all, I would like to thank Thomas Piketty for giving me a chance eight years ago. I feel privileged to have been able to learn economics under his supervision and by working with him. His vision of the discipline greatly shaped my way of doing research and will remain a constant source of inspiration. I sincerely thank Ekaterina Zhuravskaya as well, for her continuous support, for being an amazing role model and for making me discover fascinating aspects of political economy I did not know economists could study. I warmly thank Daron Acemoglu and Noam Yuchtman for agreeing to write reports on this dissertation. It goes without saying that their research has greatly influenced my own, but I would like to particularly thank them for their immense intellectual generosity, invaluable feedback and benevolence. I greatly learnt each time I had the opportunity to meet and discuss with them. I would also like to thank Mohamed Saleh who accepted to be a member of my jury, for his insightful feedback since my first year of PhD and for his inspirational work on the new economic history of the Middle East. I am truly honored to have such a jury!

Finally, I cannot thank enough Alberto Alesina, who gave an early and decisive support to my work. My year at Harvard with him was instrumental and I will always be grateful to have been able to meet and learn from such an incredible scholar and person. His passion for political economy, visible both in his research and in his dedication to guide and support students in the field, was truly inspirational and I will try my best to pay it forward in the years to come.

I am also indebted to many researchers who taught fascinating classes, took the time to give me feedback, suggested new ideas and allowed me to share my work in various seminars and institutions. In particular, I would like to thank Ayça Akarçay,

Facundo Alvaredo, Jeanet Bentzen, Tim Besley, Hamit Bozarslan, Melani Cammett, Jean-Paul Carvalho, Jad Chaaban, Melissa Dell, Ishac Diwan, Ben Enke, Kristin Fabbe, Ed Glaeser, Lisel Hintz, Timur Kuran, Sylvie Lambert, Camille Landais, Avital Livny, Ben Marx, Nathan Nunn, Vincent Pons, Pia Raffler, Nisreen Salti, Fawwaz Traboulsi, Joachim Voth, David Yang and Gabriel Zucman. Special thanks go to Emmanuel Szurek who organized the fascinating seminars series on the socio-history of Turkey at EHESS, that I have been following avidly since 2016 and from which I learnt so much.

My doctoral experience would not have been the same without the amazing PhD students I had the privilege to meet at PSE (in particular the WIL family and the R5 offices), at Berkeley (the French and Brazilian team), at LSE (the public econ office that nicely welcomed me), at Harvard (the PE tea group, the PED and Middle East Institute communities) and at AMSE (mes gâtées). Augustin Bergeron, Sophie Cottet, Caroline Le Pennec, Clémence Tricaud and Clémentine Van Effenterre provided me with crucial help during the PhD and the academic job market, and I am immensely grateful for their support and for their friendship. In particular, Clémentine, thank you for being my guardian angel, you know how much this thesis owes you.

I am also grateful to all institutions and associations that have supported my research: l'École Normale Supérieure de Cachan, EHESS, PSE, WIL, IRLE at Berkeley, LSE, Carnegie Middle East Center, Harvard, AMSE and also AALIMS, ASREC and ACES.

I finally thank my friends and family for bringing me joy, for keeping me sane (and on the right political track) and for reminding me what's important. In particular, I thank Pauline and Rafaëlle for having been there since the beginning; Mahmoud and Ghada for making Lebanon a new home in the region and for always emphasizing the beauty of our countries, cultures and histories despite all the tragedies; my brother Joud, for his wisdom, humor and sensitivity; and my parents Jalal and Rana, for inspiring and guiding me always and for their unwavering support and love.

To Alberto Alesina

SUMMARY

This PhD dissertation analyzes two dimensions of the political economy of development of the Middle East, from a contemporary and a historical perspective. The first two chapters have a particular focus on income and wealth inequality in Lebanon and in the region as a whole. The third chapter analyzes one aspect of nation-building: the role of leadership as a propaganda tool, in the context of Turkey.

In the **first Chapter**, I combine household surveys, national accounts and unique personal income tax records to produce the first estimates of the national income distribution in an Arab country, Lebanon. I find that income is extremely concentrated over the 2005-2014 period: The top 1 and 10% of the adult population received almost 25 and 55% of national income on average, placing Lebanon among the countries with the highest levels of income inequality in the world. Results are robust to various sensitivity analyses. The extreme level of inequality found questions the long-lasting narrative of the "*Lebanese economic miracle*" that showcases the country as a paragon of economic success in the Middle East. They also confirm results from a large literature, mostly in political science, that emphasizes how the Lebanese sectarian-based mode of governance has allowed the ruling elite to extract large rents on most economic activities at the expense of the majority of citizens for several decades.

In the **second Chapter**, realized in collaboration with Facundo Alvaredo and Thomas Piketty, we combine household surveys, national accounts, income tax data and wealth data in order to estimate income concentration in the Middle East for the period 1990-2016. According to our benchmark series, the Middle East appears to be the most unequal region in the world, with a top decile income share as large as 64%, compared

to 37% in Western Europe, 47% in the US and 55% in Brazil. This is due both to enormous inequality between countries (particularly between oil-rich and population-rich countries) and to large inequality within countries (which we probably under-estimate, given the limited access to proper fiscal data). We stress the importance of increasing transparency on income and wealth in the Middle East, as well as the need to develop mechanisms of regional redistribution and investment.

Finally, in the **third Chapter**, I investigate the role of leadership in constructing a national identity. I study the activities and legacy of Mustafa Kemal “Atatürk”, the founder of modern Turkey. I create a novel historical database containing information on the locations and dates of Atatürk’s propaganda visits to over a quarter of Turkish cities between 1923 and 1938. Using variation over time and across space, and information on incidental visits to districts lying along Atatürk’s road, I find that Atatürk’s visits caused an increase of 10% in the use of first names in “Pure Turkish”, the new language introduced by the state as part of its homogenizing endeavor. I argue that this measure indicates a successful diffusion of the new national identity locally. The effect is persistent, growing in magnitude up until fifteen years after the visit before disappearing. Two main channels can explain this pattern of propagation. First, the visits provided the ground for institutional reforms, as they led to the formation of local branches of Atatürk’s party. Second, the effect is stronger in districts with more nationalistic associations, higher literacy rates and where Atatürk met with local elites, suggesting that co-optation of the elite is a key driver of the effect. My findings provide new evidence on the ability of an individual leader to construct a national identity, by rallying the elite and by fostering institution building, which in turn contribute to influencing people more broadly.

RÉSUMÉ

Cette thèse de doctorat étudie deux principales dimensions de l'économie politique du développement du Moyen-Orient, en utilisant une perspective historique et contemporaine. Les deux premiers chapitres se concentrent sur la notion d'inégalité, et analysent les inégalités de patrimoine et de revenu au Liban entre 2005 et 2014 et au niveau régional entre 1990 et 2016. Le troisième chapitre se penche sur le processus de construction nationale ou "nation-building", et étudie le rôle du leader comme vecteur de propagande nationaliste dans le contexte historique turc.

Dans le **premier chapitre**, je combine de manière systématique des données d'enquête, les comptes nationaux et les seules données fiscales disponibles au Moyen-Orient pour estimer pour la première fois la répartition du revenu national dans un pays arabe : le Liban. Entre 2005 et 2014, la répartition des revenus estimée est très inégalitaire : les 1 et 10% les plus riches du pays reçoivent respectivement 25 et 55% du revenu national total, faisant du Liban l'un des pays les plus inégalitaires au monde. Les résultats sont robustes à diverses analyses de sensibilités. Ces niveaux extrêmes d'inégalités questionnent le narratif du "*miracle économique libanais*", selon lequel le Liban serait un modèle de développement économique au Moyen-Orient. Ils confirment également une large littérature, principalement en sciences politiques, décrivant les différents mécanismes par lesquels l'économie politique du pays dans son ensemble et en particulier le mode de gouvernance confessionnel ont permis à la classe politique de se maintenir au pouvoir et d'extraire des rentes colossales aux dépens de la majorité des citoyens depuis des décennies.

Dans le **deuxième chapitre**, réalisé en collaboration avec Facundo Alvaredo et Tho-

mas Piketty, nous combinons de manière systématique des données d'enquête auprès des ménages, les comptes nationaux, des données fiscales et des données sur les patrimoines provenant des recensements de la richesse des milliardaires publiés par plusieurs magazines, afin d'estimer la répartition des revenus au Moyen-Orient entre 1990 et 2016. D'après nos estimations de référence, le Moyen-Orient est la région la plus inégalitaire au monde, avec les 10% des individus les plus riches qui reçoivent près de 64% de l'ensemble du revenu national, alors qu'ils reçoivent 37% de ces revenus en Europe de l'Ouest, 47% aux Etats-Unis, et 55% au Brésil. Ce niveau extrême d'inégalités s'explique à la fois par un niveau élevé d'inégalités entre pays (en particulier entre pays richement dotés en pétrole et pays densément peuplés mais sans ressources pétrolières) et par d'importantes inégalités au sein de chaque pays, probablement encore largement sous-estimées du fait du manque de données fiscales. Ces résultats soulignent la nécessité d'augmenter la transparence et l'accès aux données dans la région, ainsi que l'importance de développer des mécanismes régionaux de redistribution et d'investissement.

Enfin, le troisième chapitre analyse le rôle que peut avoir un leader dans la construction d'une identité nationale. Plus précisément, j'étudie les activités et l'héritage de Mustafa Kemal "Atatürk", le fondateur de la Turquie moderne grâce à une nouvelle base de données historique contenant des informations sur les lieux et dates des visites officielles d'Atatürk dans plus d'un quart des villes turques entre 1923 et 1938. En utilisant la variation géographique et temporelle de ces visites, ainsi que l'information disponible sur les arrêts faits en chemin dans des villes se situant sur son itinéraire planifié, je trouve que les visites d'Atatürk ont causé une augmentation de 10% de l'usage des prénoms en "Pur turc", le nouveau langage créé par l'état afin d'homogénéiser sa population. Cette mesure permet de mesurer indirectement l'arrivée locale de la nouvelle identité nationale promue par le gouvernement kémaliste. L'effet est persistant et augmente au cours du temps jusqu'à près de 15 ans après la première visite, puis disparaît. Cette dy-

namique peut s'expliquer par deux principaux mécanismes. Tout d'abord, les visites ont permis la création d'institutions locales, puisqu'elles prédisent l'ouverture de branches du parti d'Atatürk dix ans après ses visites. Deuxièmement, l'effet est plus fort dans les districts qui avaient plus d'associations nationalistes pendant l'ère Ottomane, un plus haut taux d'alphabétisation, et où Atatürk a rencontré les élites locales, ce qui suggère que sa capacité à coopter et coordonner les élites locales a joué un rôle primordial. Ces résultats apportent de nouvelles évidences sur les mécanismes par lesquels un individu peut contribuer à la légitimation d'un nouvel ordre national.

CONTENTS

<i>Acknowledgments</i>	iii
<i>Summary</i>	vi
<i>Résumé</i>	viii
<i>General Introduction</i>	1
1. <i>Rethinking the Lebanese Economic Miracle:</i> <i>The Extreme Concentration of Income and Wealth in Lebanon, 2005-2014</i>	10
1.1 Introduction	11
1.2 Data and Methodology	16
1.2.1 First Step: Estimating a Survey Income Distribution	16
1.2.2 Second step: Fiscal Correction of the Survey Distributions	18
1.2.3 Third step: correcting for missing capital incomes	21
1.3 Results	25
1.3.1 Levels of income inequality	25
1.3.2 The distribution of economic growth	25
1.3.3 International comparisons	26
1.3.4 Wealth inequalities	27
1.3.5 "Simplified" but informative DINA: sensitivity analysis and checks	28
1.3.6 Discussion : What are the drivers of such extreme levels of inequality?	29
1.4 Conclusion	31
2. <i>Measuring Inequality in the Middle East 1990-2016: The World's Most Unequal Region?</i>	42
2.1 Introduction	43
2.2 Literature Review: Inequality Measurement in Middle East countries	46
2.3 Data sources, Concepts and Methodology	48
2.3.1 First step: constructing a household income database for the Middle East	48
2.3.2 Second step: Fiscal data correction	51
2.3.3 Third step: Missing capital income and wealth correction	54
2.4 Main Results: Extreme Concentration of Income in the Middle East	58
2.4.1 Evolution of average incomes and population in the Middle East	58
2.4.2 Extreme Level of Income Concentration in the Middle East	62

2.4.3	Analyzing the evolution of income inequality in the Middle East and other robustness checks	64
2.5	Conclusion	68
3.	<i>Charismatic Leaders and Nation-Building</i>	84
3.1	Introduction	85
3.2	Historical Background	92
3.2.1	The Turkish National Revolution, 1923-1946	92
3.2.2	“Governing with words”: the Language Reform	93
3.2.3	Spreading the words: Authoritarianism and State Propaganda	95
3.3	Data	98
3.3.1	Exposure to Atatürk during a Visit	98
3.3.2	Naming Practices	99
3.3.3	Additional data on nation-building tools	102
3.3.4	Historical and geographic covariates	102
3.4	Empirical Framework	103
3.4.1	Understanding the Logic of the Visits	103
3.4.2	Sample Restriction	104
3.4.3	Main Specifications	105
3.5	Main Results	108
3.5.1	Effects of Atatürk’s Visits on Naming practices	108
3.5.2	Addressing Identification Challenges	112
3.6	Mechanisms	116
3.6.1	Leader and institutions: Complement or Substitute?	116
3.6.2	Co-optation of local elites	119
3.6.3	An idiosyncratic effect? Comparing Atatürk’s and İnönü’s visits	122
3.6.4	Discussion: Indoctrination or Forced Assimilation?	123
3.7	Conclusion	126
	<i>Bibliography</i>	148
	<i>Appendix</i>	161
A.	<i>Appendix to Chapter 1</i>	162
A.1	Data sources and Methodology	163
A.1.1	Step 1: Household Survey Series	163
A.1.2	Step 2: Fiscal Series	164
A.1.3	Missing Capital income	165
A.2	Lebanese Income Tax	167
A.2.1	Presentation of the Personal Income Tax Law	167
A.2.2	Income definition and deductions	168
A.3	Additional Figures and Tables	170

<i>B. Appendix to Chapter 2</i>	179
B.1 Survey data - sources	180
B.2 Income and wealth distribution series	182
B.2.1 Income distribution series	182
B.2.2 Wealth distribution series	184
B.3 Additional Figures and Tables	185
<i>C. Appendix to Chapter 3</i>	193
C.1 Data sources	194
C.2 Additional Figures	212
C.3 Additional Tables	220

GENERAL INTRODUCTION

In the last decades, the Middle East has been the scene of dramatic events: wars, invasions and strengthening of authoritarian regimes. This extreme level of instability and violence contributed to describe the region as “exceptional” in terms of its political and economic development. The narrative of a Middle Eastern exceptionalism, however, suggest that the region would escape the explanatory reach of theoretical and empirical analyses (Hariri, 2015), leaving room for cultural and essentialist explanations. If the region is admittedly an outlier with regards to various key aspects of political and economic development, the historical trajectories leading to these equilibria should be analyzed for themselves, including by economists, who have relatively understudied the region.

The goal of this dissertation is to contribute to the study of the political economy of development (PED) of the Middle East, by unearthing new data sources and by applying quantitative tools to two central dimensions of the field (1) the distribution of income and (2) state capacity and nation-building.

The first part of the dissertation, which includes the two first chapters, focuses on the measurement of income inequality in Lebanon and in the region as a whole. The second part of the dissertation, which includes the third chapter, aims at understanding the causes of the extreme levels of inequality found in the first part, and studies nation-building and state institution formation in the context of Turkey. While the current chapter does not make a link between the nation-building process and the large contemporary inequality and cleavages existing in Turkey, it constitutes a first step towards our understanding of how nation-building policies were implemented and what can be their economic consequences, including on inequality levels. Chapter 3 focuses on one dimension of the Turkish nation-building endeavor: the propaganda made by

Mustafa Kemal “Atatürk”, the founder of the Turkish nation-state, to legitimize the new nation.

Part One: Income Inequality

Between 2010 and 2012, the Middle East witnessed a series of popular uprisings against the long-standing authoritarian regimes and their corrupted crony capitalist networks. The movements also highlighted deep socio-economic grievances, as demand for more social justice was among the main claims of protesters (Cammatt et al., 2015). These grievances were confirmed by a second wave in 2019. Yet, according to existing (survey-based) official estimates, income inequality in Middle Eastern countries is not particularly high by historical and international standards (Bibi and Nabli, 2009). This somewhat surprising fact has been described as the “*Arab inequality puzzle*” (World Bank, 2016).

The first two chapters are motivated by these observations and argue that the answer to the “*Arab inequality puzzle*” lies in a measurement error. Until recently, available data were insufficient to properly measure inequality. Only survey data, which notoriously suffer from top coding, underreporting, and truncations problems, were available, and often lacked information on income or were shared in tabulated form only.

The **first Chapter** combines household surveys, national accounts and unique personal income tax records to produce the first estimates of the national income distribution in an Arab country, Lebanon. I find that income is extremely concentrated over the 2005-2014 period: The top 1 and 10% of the adult population received almost 25 and 55% of national income on average, placing Lebanon among the countries with the highest levels of income inequality in the world. Results are robust to sensitivity analysis. The extreme level of inequality found questions the long-lasting narrative of the “*Lebanese economic miracle*” that showcases the country as a paragon of economic success in the Middle East. They also confirm results from a large literature that emphasizes

how the Lebanese sectarian-based mode of governance has allowed the ruling elite to extract large rents on most economic activities in the last decades at the expense of the majority of citizens.

In the **second Chapter**, realized in collaboration with Facundo Alvaredo and Thomas Piketty, we collected and referenced all existing data sources on income and wealth available in the region. In particular, we collected household surveys (micro and tabulated data), national accounts, income tax data and data on billionaires' wealth published in magazines in order to estimate income concentration in the Middle East for the period 1990-2016. According to our benchmark series, the Middle East appears to be the most unequal region in the world, with a top decile income share as large as 64%, compared to 37% in Western Europe, 47% in the US and 55% in Brazil. This is due both to enormous inequality between countries (particularly between oil-rich and population-rich countries) and to large inequality within countries (which we probably under-estimate, given the limited access to proper fiscal data). We stress the importance of increasing transparency on income and wealth in the Middle East, as well as the need to develop mechanisms of regional redistribution and investment.

Part Two: Nation-Building

Can specificities of nation-building policies explain the large levels of income inequality found in the Middle East? If nation-states have become the main political unit in the world over the last century, there is however a large variation in the success of nation building attempts, as measured by inclusive economic development, political stability and democratization ([Wimmer and Feinstein, 2010](#)). In the Middle East in particular, ethno-religious diversity and failed nation-building attempts are often described as one of the main causes of the region's political instability and relative large economic and political cleavages.

Beyond the Middle Eastern context, we still have little knowledge on which nation-building policy can foster a successful transition from a fragmented society to a nation and under which conditions. On the one hand, nation building policies, which are policies aimed at *“forming countries in which citizens feel a sufficient amount of commonality of interests, goals and preferences”* (Alesina and Reich, 2015) can bolster development, as ethnic, religious or linguistic diversity are widely understood to impede cooperation, harm the quality of governance and heighten the risk of violence and institutional breakdown (Alesina et al., 1999; Fearon and Laitin, 2003; Habyarimana et al., 2009). On the other hand, nation-building policies are often violent, amount to forced assimilation and therefore can generate backlashes, especially when imposed by a foreign power (Acemoglu et al., 2016; Dell and Querubin, 2017; Fouka, 2019; Tilly, 1993). Top-down institutional changes might also interact with local cultures in unexpected ways, and dampen or strengthen the original policy (Bau, 2021; Bisin and Verdier, 2017).

The motivation of the third chapter is to start shedding light on these issues by studying the Turkish nation-building reforms.

The Turkish context constitute a perfect lab-in-the-field to study these questions. After the defeat of the Ottoman Empire during WWI, former Ottoman military elites, led by Mustafa Kemal, secured today’s Turkish territory and created a new nation-state in 1923. They implemented radical top-down policies in order to transform the six-century old multi- ethnic, multi-religious and multi-lingual Empire into a homogeneous republic. Their goal was to establish a new identity that was secular, modern, westernized and built around the Turkish ethnicity (Zürcher, 2017). Concretely, they implemented classic homogenizing and modernizing policies, as seen in other historical contexts such as France, Italy, Iran or Egypt (Weber, 1976). They created a “Pure Turkish” language (Öztürkçe) based on new words, found in the Turkish oral tradition and folklore so that all citizens speak the same language (Szurek, 2013; Türköz, 2018). This was accompanied by a systematic repression of ethno-religious minorities and various “Turkification”

strategies, such as changing the names of minority towns and villages or banning the use of the Kurdish language (Zeydanlıoğlu, 2012). Additionally, the state used heavy propaganda. In particular during the decade that followed the creation of the nation-state, Mustafa Kemal “Atatürk” (Father Turk), the founder of modern Turkey, intensely travelled the territory and visited more than a fourth of all Turkish cities to rally citizens around his nation-building program.

The goal of the third chapter is to analyze the role and consequences of the massive propaganda effort made by Atatürk during his local visits. From 1923, the first year of the creation of the nation state, until his death in 1938, Atatürk visited more than a quarter of all Turkish cities to rally citizens around the state’s new ideology. Using a novel historical database, I assess the impact of his visits on national identity and support for the nation-building reforms. Exploiting time and geographic variation in Atatürk’s visits in a difference-in-differences strategy, I find that visited districts are more likely to use first names in “Pure Turkish”, the new language introduced by the state as part of its homogenizing endeavor. The effect is persistent and its magnitude growing over time. The effect is larger in places with former nationalistic associations, higher literacy rates and where he met with local elites, suggesting that co-optation of the elite is a key mechanism. Moreover, I find that districts that were exposed to Atatürk were more likely to form a local cultural branch of Atatürk’s party, which, in turn, also contributed to the diffusion of the new names, suggesting that the leader and institutions can act as complement. This paper provides the first empirical evidence on the ability of an individual leader to construct a national identity, by rallying the elite and by fostering institution building. The fact that Atatürk’s effect was mostly successful among a segment of the population, the elite, might have provided the ground for large societal divides (including between the masses and the elite). Whether the Turkish nation-building process contributed to create the large economic and political cleavages we still observe today in Turkey will be the object of future research.

Data and Methodology

To analyze these two dimensions of the PED of the region, this dissertation has a data-intensive approach and builds upon several empirical methods.

The first chapters are mostly descriptive and collect all existing sources on income and wealth for the region between 1990-2016: household surveys (individual level and tabulated data), national accounts, billionaires rich list, government finance reports, and the first and only micro tax records available in the region, in Lebanon. This data collection effort is a key contribution of the papers. Then I combine these sources in a systematic manner following the “*Distributional National Account*” methodology developed by [Alvaredo et al. \(2016\)](#) to produce the first micro distributions of national and regional income in Lebanon and at the regional level that match macro aggregates. In particular, novel Pareto interpolation techniques developed by [Blanchet et al. \(2022\)](#) allow me to make use of survey tabulated data previously not exploitable.

The last chapter is based on the collection and digitization of a large number of archival records and uses tools of causal inference. In particular, I collected new sources on all Atatürk’s visits and on the activities he conducted locally; historical road and railway networks; information on past Ottoman nationalistic associations, the “Turkish Hearths” and on the “People’s Houses”, associations linked to Atatürk’s main political party (and only ruling party between 1923-1946) the Republican People’s Party; all historical newspapers and booklets published in the 1930s to diffuse the new words in “Pure Turkish”, the biographies of all members of the Turkish Grand national Assembly between 1920 and 2010.¹ To estimate the impact of Atatürk’s visits, I use a difference-in-difference approach exploiting variation over time and across space in Atatürk’s visits and leveraging information on incidental visits to district lying along Atatürk’s road.

Overall, this dissertation contributes to the collection and analysis of new data sources

1. All sources are described in more detail in Chapter 3.

in the Middle East, a region where data transparency and accessibility is arguably of lower quality (Bibi and Nabli, 2009).²

Overall contribution

The political economy of development of the Middle East is a large research field, which has however been relatively less invested by economists. In particular, little work has been done on income inequality and nation-building policies in the region within economics (Blaydes, 2017; Cammett and Diwan, 2019; Cammett et al., 2015; Diwan et al., 2019; Hertog, 2017). On the inequality front, existing scholarship has mostly tried to explain the “Long Divergence” that is the lack of democratization and relative underdevelopment of the region compared to the West (Kuran, 2010).³ It has been less focused on analyzing inequality levels and determinants within the region or even within its countries. Besides, research on nation-building and nationalism has been extremely euro-centric and tends to posit stronger civic identity as “better” and incompatible with ethno-religious ones (Mylonas and Tudor, 2021). There is relatively less work on specific nation-building processes from the region.

This dissertation therefore makes several key contributions. By producing new data on the level and the dynamics of income inequality, it contributes to correct a previously pervasive but false narrative according to which income inequality was not that high in the Middle East. Besides, if most countries in the region are classified as middle-income country, our results suggests that these relatively high levels of national income per capita actually hide a very polarized socio-economic structure, with extremely poor and extremely rich individuals, whose income levels are broadly comparable to their counterparts in high-income countries. This underlines the importance of studying and

2. This assertion, however, does not apply to historical sources: the region’s archives are on the contrary relatively quite rich as emphasized by Saleh (2017).

3. This strand of literature has mostly been theoretical and/or focused on analyzing how specific and mainly “Islamic” institutions, explains the “Long divergence” (Bisin et al., 2021; Blaydes and Chaney, 2013; Chaney, 2016; Kuran, 2004; Rubin, 2017). For a more detailed review of this literature, see Saleh (2017).

tackling poverty in the region (Assouad et al., 2018).⁴

Second, this dissertation contributes to the literature on nation-building. Existing works have focused on understanding why some states start implementing nation-building policies (Alesina et al., 2021; Alesina and Reich, 2015; Bandiera et al., 2019) and on analyzing several potential determinants of nation-building such as propaganda (Blouin and Mukand, 2019), military action (Dell and Querubin, 2017), education (Bazzi et al., 2018; Blanc and Kubo, 2021), population resettlement programs (Bazzi et al., 2019) or sport (Depetris-Chauvin et al., 2020). I focus on leadership, which this literature has not systematically examined with quantitative data yet and provide new evidence on the channels through a national identity is constructed. I also provide a novel measure of national identity, that can be measured at the local level and studied through time: the adoption of first names in “Pure Turkish”, the new language introduced by the state as part of its homogenizing endeavor.

Finally, the third chapter also contributes to the historiography of Kemalism and of the Turkish nation-building era. There is an abundant literature on Kemalist reforms and their impact on the new Turkish national identity (Aymes et al., 2014; Szurek, 2018). A first historiography argues that there is a strong continuity between Kemalist elites’ goals and real socio-political transformations. This body of works stays close to the official and top-down narrative the state produced about itself, according to which reforms were extremely successful in quickly spreading a new Turkish identity and in imposing new socio-cultural norms to the periphery (Berkes, 1964; Lewis, 1974). This narrative has nevertheless been criticized by a more recent historiography that questions the view of a “Kemalist steamroller” and of a “total transformation” of Turkish citizens by the revolution (Szurek, 2018) This literature underlines that the new state did not have the financial means to reach the periphery and in particular the countryside, where 80% of the pop-

4. The fact that most countries in the region are classified as middle-income countries and not low-income countries might partly explain why the region is relatively understudied by development economists.

ulation was living at the time (Pamuk, 2018; Tuna, 2018). Besides, even when the central power was present locally, some places developed "day-to-day forms of resistance" and ways to negotiate with the central state (Brockett, 2011; Lamprou, 2015; Sakalli, 2019; Yilmaz, 2013). The paper contributes to this literature by showing that Kemal did have an average positive impact, in urban areas he visited and only on a segment of the population only, the elite.

1. RETHINKING THE LEBANESE ECONOMIC MIRACLE:
THE EXTREME CONCENTRATION OF INCOME AND WEALTH IN
LEBANON, 2005-2014

1.1 Introduction

How unequal are Arab countries? In the last decades, the Middle East has been characterized by an extreme predominance of violence, a rise of armed non-state actors and a consolidation of authoritarianism. In this context, it is important for policy makers and scholars alike to understand whether this extreme political instability is linked to the underlying distribution of economic resources and power. The popular uprisings that recently shook the region - from the 2011 "Arab Spring" to the 2019 protests - suggest that economic inequality might indeed be quite large. More social justice was among the main demands of the protesters, along with more civil and political rights. Yet, existing studies and official inequality estimates in the region suggest that income and wealth inequality levels are not that high by international standards. In 2011, the Lebanese or Egyptian official GINI indexes were below 0.35 for example, meaning that both countries were as egalitarian as the most egalitarian countries in history such as Scandinavian countries in the 1980s.¹ This somewhat surprising fact has been coined "*the Enigma of Inequality*" (UNDP, 2002) or the "*Arab Inequality Puzzle*" (World Bank, 2016)

The goal of this paper is to provide an answer to this puzzle, by studying the Lebanese case. To do so, I collected unique and novel fiscal micro-data for the 2005-2014 period from the Lebanese Ministry of Finance. I combine them with existing survey data, national accounts, billionaires' wealth data and government finance reports in a systematic manner in order to produce the first estimates of the national income distribution in a Middle Eastern country. I follow the standardized methodology of the "*Distributional National Accounts*", developed by (Alvaredo et al., 2016) and which was first applied to the US (Piketty et al., 2016). This method proposes to distribute total national income across individual adults and has recently been applied to a growing number of coun-

1. Hlasny and Verme (2018) and the World Bank Povcalnet Database.

tries, as reviewed in the World Inequality Report 2018 (Alvaredo et al., 2018). This study is the first to apply it to a Middle Eastern country, where data quality and transparency is arguably of lower quality compared to other world regions.²

The key feature of this methodology is to use fiscal data to correct survey-based estimates of inequality, as it is now widely acknowledged that surveys fail to capture the top tail of the income distribution. The problem is particularly acute in developing countries and in regions of the world where inequality might be particularly high.³ By linking the corrected income distribution to national accounts, this method produces series that are consistent with macroeconomic figures, homogeneous over time and comparable across countries. It also allows researchers to look at the entire distribution and to study the distribution of growth among all income groups.

I find that the top 1 and 10 percent of the adult population receive almost 25 and 55 percent of total national income, which places Lebanon among the countries with the highest levels of income inequality in the world, alongside Brazil, Russia, South Africa and the United States (Alvaredo et al., 2018). The Lebanese income distribution appears to be extremely polarized. The top 10 percent richest individuals receives five times as much as the bottom 50 percent of the population. The middle 40 percent of the distribution, which broadly speaking represents the middle class, is left with close to 30 percent of the total national income, which is far less than the top 10 percent. This is quite different from what we observe in Europe or in the United States, where the middle class receives more or about the same income share as the richest 10 percent over the same period. My results are still subject to major shortcomings due to data limitations, which I describe in detail in the paper. I implement various robustness checks and produce

2. See Bibi and Nabli (2009) for a review of existing data sources in the region and an assessment of their quality, and the "Inequality Transparency Index" at the WID.

3. See Assouad et al. (2018) for a detailed methodological discussion on the issue.

variant series for on each hypothesis made during the estimation procedure. In order to get an overall "confidence interval" of the inequality estimates, I replicate my procedure and choose the most or the least conservative assumptions.⁴ I find that my benchmark results are subject to a large uncertainty, with almost 10 percentage point of variation between the lower and upper bound estimates. Nevertheless, they remain robust: In all specifications, even in the most conservative one, the 10 percent richest adults receive more than 50 percent of total national income.

This paper makes two main contributions. First, it provides the only reliable estimates of the national income distribution in the Middle East, and therefore contributes to answer the so-called "*Arab Inequality Puzzle*".⁵ In the case of Lebanon, the puzzle went as far as creating an opposite narrative, according to which Lebanon was the "Switzerland of the Middle East" and a paragon of economic success in the Middle East. According to this widespread narrative, sometimes coined the "*Lebanese economic miracle*", the country would economically perform better than its neighbors, despite numerous political shocks, and ensure a relatively high level of income per capita to its citizens.⁶ This paper, by creating new data on inequality can inform public debates and shed new light on recent economic and political developments in the country, including the 2019 "October Revolution".⁷

4. I do so even when such assumptions are not empirically credible. This nevertheless enables me to produce a decision tree that highlights the lower and upper bound estimates of my results.

5. This study is the first to use fiscal data to correct the top of the survey income distribution in an Arab country. To my knowledge, the only other study correcting official survey estimates in the region is [van der Weide et al. \(2016\)](#), which uses housing price data to estimate the top tail of the income distribution in Egypt. They find that inequality levels are way higher than existing survey-based estimates. They however cannot recover the full distribution of national income in the absence of administrative fiscal data. Other studies have investigated the roots of the puzzle such as [Devarajan and Ianchovichina \(2018\)](#) who study complementary sources of dissatisfaction including dissatisfaction with the quality of public services, the shortage of formal-sector jobs, and corruption.

6. The narrative lasted as there was actually no estimates of income inequality in the country before this study. The last income share figures published for Lebanon date back to 1960 ([Ministry of Planning, 1960](#)). The only recent study available is based on information on consumption from survey data and focuses on poverty ([Laithy et al., 2008](#)).

7. As other protests in the region, more social justice was among the main demands of the Lebanese who took the streets in October, 2019. It is significant that the trigger of the uprising was a new tax on

The second contribution is methodological. Lebanon is a case where data quality is particularly low, despite the existence of micro-fiscal data. Yet, this study shows that even when data transparency and quality are very limited, it is possible to reach conclusion on inequality levels, as long as all hypotheses and data limitations are well-stated and their impact on the final series highlighted. I propose to complement the standardized DINA methodology by a systematic sensitivity analysis that consists in clarifying each assumption made during the procedure and how their combination impacts the level of inequality.

Related Literature This paper adds to the literature on the measurement of poverty and inequality in developing countries. There has recently been a growing interest for the study of income distribution, after a relative hiatus since Kuznet's seminal work in 1955 (Kuznets, 1955). A first wave of this literature has constructed top income shares time series over the long run for more than twenty countries using fiscal data (Atkinson and Piketty, 2007, 2010; Atkinson et al., 2011). Recently, this literature has attempted to estimate the full distribution of national income, using fiscal data combined systematically with survey data and national accounts, in order to estimate "Distributional National Accounts".⁸ These series follow a standardized methodology, described in (Alvaredo et al., 2016), that however needs to be adjusted depending on the data quality and availability in each specific country. When exhaustive micro-data are available (as in the US or in France), it is possible to derive "sophisticated" and precise DINA (Garbinti et al., 2018; Piketty et al., 2016). However, when data sources are limited, as in China, Russia, Brazil or in the Middle East, one needs to make more assumptions to derive "simplified" DINA (Assouad et al., 2018; Morgan, 2017; Novokmet et al., 2017; Piketty et al., 2017). My results on Lebanon belongs to the second category, and offers "simpli-

WhatsApp and other mobile applications, adding to a long list of austerity measures announced earlier in the year and which disproportionately affected the most vulnerable among the population.

8. All country specific studies and data can be found online at <http://wid.world>.

fied DINA” for the first time in an Arab country. While estimates are highly uncertain, the results on inequality are robust. This demonstrates that the relative high income per capita in Lebanon, as in other countries in the region such as Egypt or Jordan, might be driven by a rich and small group of people at the top, and hide high poverty levels.⁹

Second, this paper contributes to a literature on crony capitalism and its distributional consequences. The results provide quantitative support to a large literature on the political economy of the Middle East as a whole and of Lebanon in particular. This literature has documented various mechanisms by which institutions contribute to develop a crony capitalism in most countries of the region, which feeds sectarian clientelism, blurs the lines between public and private capital and fosters rent seeking behaviors from the highly connected political and business elites (Baumann, 2017; Chaaban, 2016; Corm, 2005; Diwan and Haidar, 2021; Diwan et al., 2019; Gaspard, 2004; Rijkers et al., 2017; Traboulsi, 2012).¹⁰

The remainder of the paper is organized as follows. In Section 1.2, I describe the data sources and methodology used. Section 1.3 presents the results on the levels of income inequality in Lebanon between 2005 and 2014 and compares them to other countries. Section 1.4 concludes.

9. The fact that most countries in the region are officially classified as “middle income countries” might partly explain why the region is relatively understudied in development economics. It is striking for example that the Middle Eastern J-PAL office opened in 2020.

10. Alvaredo et al. (2019) builds on the results for Lebanon to estimate inequality statistics at the regional level between 1990 and 2016.

1.2 *Data and Methodology*

This paper uses five main data sources: household surveys, national accounts, public finance reports, wealth rankings and importantly newly available fiscal micro data. I combine these sources in a systematic manner, following the “Distributional National Accounts” (DINA) guidelines (Alvaredo et al., 2016). This standardized methodology uses the same data-sources for all countries in order to produce estimates of the distribution of national income comparable across time and space. It broadly consists of three main steps: (1) estimating the country’s income distribution using household survey data, (2) correcting the income levels at the top of the survey distribution with fiscal data and Pareto-Interpolation, (3) adjusting the final distribution to account for missing non-fiscal and tax-exempt incomes, using national accounts and rich lists published by magazines. The approach adopted for Lebanon follows the same structure, with some adaptations due to the data format and quality described in the following sections.¹¹

1.2.1 *First Step: Estimating a Survey Income Distribution*

Lebanese survey data are scarce. Three nationally representative surveys have been undertaken recently, in 1997, 2004 and 2007.¹² The micro-data are difficult to access: The Lebanese statistical institute, the Lebanese Central Administration of Statistics (CAS) is not allowed to share data with researchers. Only Laithy et al. (2008) got access to micro-data on consumption and could estimate the bottom of the consumption distribution.¹³ The only other existing figures on the entire income distribution date back from the first

11. More details on the methodology is available in the Appendix below. An online appendix that includes all raw data sources and computer codes is available at <https://wid.world/country/lebanon/>.

12. See Table 2.1 p29 in World Bank, 2016, for a review of existing survey-based studies.

13. They document that nearly 8 percent of the population, that is 300,000 individuals, live under conditions of “extreme poverty” (less than US\$ 2.40 per day) and are not able to meet most basic food and non-food needs. They however find a relatively low Gini coefficient of 0.37 for the consumption distribution.

nationally representative survey conducted in 1960.¹⁴ I unfortunately could not access micro-data on income. I therefore used two tables published in official report by the CAS and which indicate the household frequencies for thirteen income groups, for 2005 and 2007 (before and after the 2006 war). Using the generalized Pareto interpolation techniques developed by [Blanchet et al. \(2022\)](#), I estimate the full distribution of income expressed in generalized percentiles for the two years.¹⁵

Limits Four main limitations should be stressed. The first one is related to the unit of observation. In order to follow the DINA guidelines, I take the adult individual (i.e. aged 20 and more) as the basic unit. However, there is no information on the average number of adults in each household, by income bracket. I therefore take the average number of adults per household at the national level, and assume that income is equally split between adult household members. I apply the same adults/children ratio to all brackets if high earners have fewer children than average, inequality is slightly underestimated. Second, the survey tabulations do not provide detailed information on income categories. We therefore do not know which income type is included in the overall "household income" variable and how the income concept captured in the survey data matches the one from the fiscal data and from the national accounts. The third issue concerns the years without data. I only use the 2007 survey data. More specifically, I use the tabulation titled "before the war" to estimate the 2005 and 2006 distributions and the tabulation "after the war" for the following years. I then anchor all income distribution to the relevant annual average income, that is for every year, I

14. The 1960 study shows large income disparities, with the richest 4 percent receiving 32 percent of total income while the following 14 and 32 percent have respectively 28 and 22 percent. The remaining half of the population is left with 18 percent of the national income, including 2 percent for the poorest 9 percent ([Ministry of Planning, 1960](#)).

15. Generalized percentiles (or g-percentiles) are 127 income groups along the income distribution: 99 for the bottom 99 percentiles, 9 for the bottom 9 tenth-of-percentiles of the top percentile, 9 for the bottom 9 one-hundredth-of-percentiles of the top tenth-of-percentile, and 10 for the 10 one-thousandth-of-percentile of the top one-hundredth- of-percentile. The interpolation code is available at <http://wid.world/gpinter/>. This method allows the estimation of income distribution using tables with even few income groups.

proportionally upgrade income levels for all percentiles so that per adult average income coincides with per adult average national income observed in the WID macroeconomic database. By construction this has no impact on income shares (inequality levels are the same for the 2005-2014 period). This means that I cannot draw robust conclusions on the evolution of inequality, but only on the levels. In particular, the effect of the large Syrian refugees influx after 2011 on inequality is not taken into account (except through their aggregate effect on average income).¹⁶ Finally, the ratio between total survey income and national income equals 37 percent in Lebanon, which is quite low. In many developing countries and in particular in regions with extreme levels of inequality, this ratio typically varies between 40%-50% (Assouad et al., 2018). Lebanon has also a relatively lower coverage compared to other Middle Eastern countries.¹⁷

1.2.2 *Second step: Fiscal Correction of the Survey Distributions*

The second step consists in correcting the top of the survey distribution using fiscal data. We now know that inequality statistics based on surveys are seriously downward biased, due to under-reporting, truncations and top coding problems at the top (Bourguignon and Morrisson, 2002; Burkhauser et al., 2016; Dowrick and Akmal, 2005; Jordá and Niño-Zarazúa, 2019; Lakner and Milanovic, 2013). Besides, survey data often only covers a small share of total national income, especially in developing countries. To the extent that this missing income generally accrues to relatively small groups of the population, this implies that survey-based statistics may severely underestimate income inequality. To tackle this issue, some studies attribute all missing income to the top 10 percent income recipients, or use Pareto-type imputations to distribute the missing income (Burkhauser et al., 2016; Lakner and Milanovic, 2013). My strategy is to merge the survey and fiscal data using the “generalized Pareto interpolation” method developed

16. Household tabulations are also available in 2004. See section 1.3.5 for robustness checks of the impact of the choice of the survey year.

17. See Table 2, p6 in ?, which displays the average ratio (total survey income)/(national income) for all countries in the Middle East with survey data.

by [Blanchet et al. \(2022\)](#). This strategy has the advantage to rely on additional data and on better estimation techniques for the very top of the income distribution. It is particularly suitable to the Lebanese case, as fiscal data are of much quality than the survey data, which is rarely the case in other contexts. In the rest of this section, I briefly present the Lebanese personal income tax data and describe the correction procedure.

The Lebanese micro-fiscal data

The Lebanese Personal Income Tax (PIT) created in 1959 is a schedular, progressive and individual tax which taxes separately: (1) some business incomes (profits made by self-employed individuals, partners in partnerships and individuals in small corporations) at marginal rates ranging from 4 to 21 percent, (2) labor income (salaries, wages, bonuses, allowances, life annuities, pension payments, and other benefits in cash and kind) at rates ranging from 2 to 20 percent and, finally, (3) rental revenues from built property at rates ranging from 4 to 14 percent. Next to the personal income tax, incomes from movable capital (dividends incomes, board member appropriations from profits and interest incomes, including interest on bonds and treasury bills) are taxed at flat rates. I do not observe those incomes, as I could only access tax records of the PIT. The database shared by the Ministry of Finance is an unbalanced panel. Each observation corresponds to the annual declaration of one taxpayer and the three sources of income listed above are reported separately. For business income and wages, gross income (before any deduction and gross of expenses) and taxable income (after deductions of charges and benefits). For rental revenues, only taxable income is reported. The data is reliable for the top 1 percent of the adult population, although it covers a greater share of the adult population.¹⁸

18. The database covers up to 15 percent of adult individuals in some years. However, due to the schedular form of the tax, individuals in lower income groups, receiving low wages, are included in the database even if they do not belong to the top 1 percent income group.

Correcting the top of the distribution

Given that the tax records only provide information on the total gross income of an individual or on its taxable income, I need to make assumptions to obtain the actual individual fiscal income (pre-tax, pre-deductions fiscal income but net of expenses). In my benchmark series, I assume that taxable income equals 80 percent of total fiscal income.¹⁹ Next, I consider that the survey distribution estimated in step 1 is reliable for the bottom 80 percent of the distribution (below the 80th percentile, $p_1 = 0.8$) and that the fiscal data are reliable for the 99th percentile and above ($p_2 = 0.99$). In order to link the two distributions, I assume that the quantile ratio upgrade factor $f(p)$ rises piecewise-linearly from $f(p_1) = 1$ to the observed fiscal/survey ratio between p_1 and p_2 , $f(p_2)$, so as to generate a smooth and convex Pareto curve Blanchet et al. (2022). I then apply generalized Pareto interpolation techniques to the corrected tabulations to obtain the full distribution of fiscal income among equal-split adults, by g-percentiles, between 2005 and 2014.²⁰ While recent research has shown that survey and tax data seem to start diverging at $p = 0.90$, I choose to use the survey data up to the 80th percentile as it yields to more conservative estimates (Yonzan et al., 2021).²¹

Limits Most capital incomes (imputed rental revenues of persons living in their own dwelling, dividends, interests, or profits made by individuals in limited partnerships, joint stock or limited liability companies) are not reported in the fiscal data. More generally the personal tax records miss income which evades from taxation and income from the informal sector. The third step of the correction procedure, presented in the following section 1.2.3, partially accounts for these issues.

19. Total taxable income is the sum of taxable business income, wages and housing rents. See Section 1.3.5 for robustness checks on the impact of these two assumptions.

20. I also provide several variants based upon different piecewise-linear profiles for the upgrade factor between $f(p_1)$ and $f(p_2)$, and the share of the total distribution covered by the survey data (see Section 1.3.5).

21. This study is based on data from the United States, Germany, and France. See section A.1.2 in Appendix for a more detailed discussion on this step, and for sensitivity analysis.

1.2.3 *Third step: correcting for missing capital incomes*

The last and final step of my estimation procedure corrects for missing capital incomes, that is tax-exempt and non-reported capital incomes. This third step itself contains three main sub-steps described in the rest of this section.

Estimating and reallocating the amount of income missing

First, I estimate the size of the missing capital income in terms of national income. A natural way to recover the macroeconomic amount of capital income not taxed under the PIT data is to look at national accounts. However, in Lebanon, national accounts are of very poor quality and are not disaggregated enough. I therefore look at Public Finance reports, which give for each year the amount of tax revenues collected for each tax.²² I recover the missing amount by dividing the revenues collected from the different income sources by the corresponding tax rate in force in the legislation. I find that non-reported and tax-exempt capital incomes represent approximately 20 percent of national income. Then, to estimate the final distribution of total personal income (y_p), the sum of fiscal income (y_f) and missing income (y_m), I first assume that y_m follows the same distribution as wealth below for the estimation of the wealth distribution. As for the correlation structure between y_f and y_m , I use the family of Gumbel copulas, with Gumbel parameter $\theta = 2$ (Novokmet et al., 2017; Piketty et al., 2017).²³ In order to compute the joint distribution of fiscal and non fiscal income, I therefore need to estimate the distribution of wealth in Lebanon as I assume that (y_m) follows the same distribution.

22. The Public Finance Reports are available online, on the website of the Lebanese Ministry of Finance, <http://www.finance.gov.lb/en-us/Finance/Rep-Pub/DRI-MOF/PFR>

23. See the detailed computations in the Appendix and section 1.3.5 for variant series depending on the total amount of missing capital income reallocated and the Gumbel parameter chosen.

Estimating the Lebanese Wealth Distribution

Wealth data are scarcer than income data in Lebanon. Only billionaires' lists, published by Forbes and the magazine Arabian Business, are available. I nevertheless take advantage of these sources of information, and compute the ratio of billionaires' wealth to national income. I use this as a proxy to compare the "weight" of billionaires in various countries' economies. As displayed in Figure 1.1, billionaires' wealth represents 30 percent of total national income on average over 1990-2016, surpassing by far what we observe in other countries using the same data. The conclusion is similar if we look at the average between 1990 and 2005 or 2005 and 2016. This relative important "size" or weight of billionaires' wealth, expressed as a function of national income, suggests that wealth is more concentrated in Lebanon. Then, given that there is no survey on wealth for Lebanon, I proceed as follows. I compute an average standardized distribution of wealth for the US, France and China, for which we have reliable estimates of wealth inequality. More precisely, I divide all thresholds and bracket averages for all percentiles by the average wealth, and compute the arithmetic average for the three countries.²⁴ Variations across countries and over time in these standardized wealth distributions mostly happen above $p_0=0.99$, that is, for the bottom 99 percent of the distribution, average wealth is relatively stable. Therefore, I take the same normalized distribution for Lebanon below $p_0=0.99$ as the average US-France-China normalized distribution, hereby assuming that wealth is at least as concentrated in Lebanon as in countries with available data. To estimate the Lebanese average wealth, necessary to derive the final wealth distribution, I also compute an annual average wealth/income ratio for all countries with available data on [WID.world](#). I find that, on average, wealth represent at least 300 percent of total national income in countries for which we have data. I therefore consider that the average wealth in Lebanon for a given year t equals $\frac{Wealth/Income_{wid} \times NationalIncome_{Leb,t}}{AdultPopulation_{Leb,t}}$. Once the average normalized wealth distribution computed and adjusted to match the

24. I take the data from [WID.world](#), using the "wid" STATA command.

estimate average wealth in Lebanon, I need to take into account the weight of Lebanese billionaires at the top. This is not obvious, as I need to link the 99th percentile to the few billionaires at the very top, making also assumption on their family size (to know how many individuals benefits from the Lebanese billionaires' wealth). I therefore need to make an assumption about the average number n of adults per billionaire family (sometimes Forbes includes very large family groups in the same billionaire family; sometime it is just one individual or one married couple) and on the correction profiles to link the normalized wealth distribution until the 99th percentiles towards the billionaires.

To summarize, this procedure consists in assuming that the Lebanese total wealth and wealth distribution are similar to what we observe in other countries on average, and to correct the top of the distribution by taking into account the relative importance of the Lebanese billionaires' wealth.²⁵

Adjusting the final series to macroeconomic average income

Figure 1.2 shows the share of the total national income covered by each data source. At the end of the three corrections, there are still 30 percent of the total national income missing, which is quite large but similar to other developing countries with relatively high levels of inequality.²⁶ The remaining 30 percent contains a combination of incomes that evaded taxation and incomes made in the informal sector. The latter are partly taken into account in the bottom on the distribution by the survey data, so a large share of this 30 percent should probably accrue to the top groups. I nevertheless chose in my benchmark to proportionally upgrade all income levels at all percentiles so that per adult average income always coincides with per adult average national income (therefore keeping the income distribution and shares constant).²⁷

25. See section 1.3.5 for robustness check on the assumption made to estimate the Lebanese wealth inequality. This methodology is also used for other Middle Eastern countries in [Alvaredo et al. \(2019\)](#) and for Russia [Novokmet et al. \(2017\)](#).

26. See Figure 1 in [Assouad et al. \(2018\)](#)

27. An alternative would be to allocate proportionally the 30 percent toward the bottom 50 percent income group. This implicitly assumes that this amounts mostly come from the informal sector and/or goes

Limits My estimates of wealth inequality used in this step are highly uncertain. I simply assume that Lebanon should have a total amount and a concentration of wealth that are at least as high as what we observe in other countries. The only data used are the billionaires' worth list, which are particularly fragile and volatile in Lebanon (only 7 billionaires are reported, and some years do not have data). Using this data source to identify a trend in wealth concentration is impossible. Nevertheless, the stable and high concentration revealed in the rich lists reflects something real about the Lebanese wealth distribution and the method might at least give a good first approximation of the concentration of wealth in the country. Given the political economy of the country, this assumption is credible and the high levels of inequality found not so surprising. The Lebanese economy lies mostly on the banking and real-estate sectors. The Bank Secrecy Law of 1956 made Lebanon the "Switzerland of the Middle East". Most importantly, the country is characterized by a crony capitalism where political elites and business elites are highly connected (see the discussion section 1.3.6). They share between themselves the main sectors of activities and companies of the country and extract large rents on any income generated in the country (Chaaban, 2016; Diwan and Haidar, 2021; Gaspard, 2004). Reassuringly, estimates on wealth inequality are only used in the third step, which has a limited impact on the final income distribution, compared to the fiscal correction (see Section 1.3.5 for the decomposition of the effect of each correction).

to the poorest, which is not credible. Doing so nevertheless does not affect much the main conclusions.

1.3 Results

1.3.1 Levels of income inequality

The main results of the paper are summarized in Figure 1.3. Income is extremely concentrated in Lebanon, with the richest 10 and 1 percent adults accounting for almost 55 and 25 percent of total national income, on average throughout the period. In contrast, the bottom 50 percent of the Lebanese population is left with approximately half of what is accruing to the top 1 percent. Figure 1.4 gives a sense of the extent of the concentration: the top 0.1 percent of the adult population, that is approximately 3000 individuals receives approximately the same amount of national income as the bottom 50 percent, that is 1,5 million individuals. Finally, the middle 40 percent of the Lebanese adult population receives one third of the total national income. My estimates are consistent with the high levels of poverty reported in [Laithy et al. \(2008\)](#), although we do not use the same welfare concept and unit of observation. I find a higher Gini coefficient, mostly due to the fiscal correction. It should be emphasized that given the lack of yearly survey data, results on the evolution and dynamics of inequality levels are uncertain. Besides, the Lebanese income distribution appears to be extremely polarized. The middle 40 percent of the distribution, which broadly speaking represents the middle class, is left with close to 30 percent of the total national income, which is far less than the top 10 percent. This is quite different from what we observe in Europe or in the United States, where the middle class receives more or about the same income share as the richest 10 percent over the same period (see Figure 1.5).

1.3.2 The distribution of economic growth

Between 2005 and 2014, real national income increased steadily, with a cumulated growth rate of almost 50 percent (Figure 1.6). However, if we look at the per adult national income, it follows a bell-shaped curve, increasing between 2005 and 2010 and then

decreasing due a sharp population growth of 50 percent, mostly following the major inflow of Syrian refugees. We therefore observe a slight impoverishment of the Lebanese population after 2011, which lost on average 2 percent of its yearly real income. The series computed in this paper allow me to go further and to determine which income groups did or did not benefit from growth. Figure 1.7 shows that the bottom 90 percent of the adult population experiences a negative growth, far below the average, while the top 10 percent enjoyed very large growth rates.²⁸ In order to understand the driving forces behind these high growth rates at the top, I examine the respective role of business income, labor income and rental revenues using the fiscal micro-data. Figure 1.8 decomposes top groups by income categories for the years 2005 and 2014. This figure should be interpreted with caution as it only captures capital incomes subject to the PIT tax.²⁹ Several remarks can be made. First, the negative growth rate of the top 0.01 percent seems to come from a sharp decline in rental revenues over the period, which translated into an increase in the share of wages. A first explanation for this is the major property destructions that happened during the Israeli war.³⁰ However, as early as 2007, a massive reconstruction effort was made and demand on housing kept increasing while real-estate prices and rental income skyrocketed. The variation we observe at the very top may simply reflect a change in tax evasion behaviors due to the political instability that began in 2005.

1.3.3 *International comparisons*

Figure 1.9 compares the top 10 and 1 percent income share in Lebanon with series for Brazil, China, France, Russia and the United States. The conclusion is clear: Lebanon has one of the highest records of income concentration in the world. Tables 1.1 and

28. Except for the top 0.001 percent (that is between 25 and 37 adults over the period), for which the rate becomes negative again.

29. This probably means that the micro-files might not exactly represent the top 1 percent but rather fractions of individuals in top groups (perhaps within the top 10 percent).

30. The Israeli war indeed damaged more than 210,000 housings and destroyed 25,000, leaving more than 300,000 people homeless [Verdeil \(2006\)](#).

1.2 present the income thresholds and averages within the different income groups, in 2016 Euro PPP in Lebanon and in other regions of the world. To be among the 1 percent richest Lebanese, one needs to make at least 123,651 Euro per year in 2016, for an average income of 335,930 Euro, levels comparable to Western Europe. The magnitude of concentration however increases drastically within top groups, with an average income for the top 0.1 percent of 1,593,622 Euro. To get a sense of the skewness of the Lebanese distribution, it is interesting to compare the average income within each group in Lebanon and in Western Europe. Until the top 1 percent, the average income is systematically smaller in Lebanon, representing 40 percent of the corresponding average in Western Europe for the bottom 50 percent and 90 percent for the top 1 percent. Within top groups, the ratio reverses to reach 140 percent within the top 0.01 percent and even 190 percent within the top 0.001 percent. In other words, in Lebanon the richest are as rich or richer than their counterparts in Western Europe, while the poorest are way poorer. The average income of individuals at the very top of the distribution in Lebanon is broadly comparable to average levels observed in Brazil or South Africa, other extremely unequal countries. Lebanon is therefore characterized by a polarized income structure, without a broad "middle class" comparable in size to the one in high-income countries.

1.3.4 *Wealth inequalities*

Figure 1.10 reports statistics on the average concentration of wealth for the 1990-2016 period, obtained using data from the annual Forbes and Arabian business rankings that cover the wealthiest Lebanese individuals.³¹ According to my benchmark estimates, wealth is on average extremely concentrated with the top 10 and 1 percent of the Lebanese adult population gathering almost 45 and 70 percent of total personal

31. The only other existing estimates of wealth inequality in Lebanon are the one by Davies et al. (2010-2016), which also use rich list and Pareto interpolation techniques. Unfortunately, as emphasized in Novokmet et al. (2017), their estimation technique is not explicit (one cannot replicate their results, and there is no online code available).

wealth respectively.³² These levels are substantially higher than in China and France and slightly higher than in Russia and the United States in the recent period (Figure 1.10).

1.3.5 *"Simplified" but informative DINA: sensitivity analysis and checks*

The previous sections underline that the series derived in this paper are subject to high uncertainty. In most developing countries, one can only derive "simplified" DINA, given the difficulty to access administrative data and to track income in a satisfactory manner. This seems to be particularly true in regions considered to be extremely unequal.³³ I argue that despite data limitations, it is still important to attempt to measure income inequality in these regions, as long and only if the assumptions made and their implications are clearly stated. Figures A.1 to A.7 show the impact of each hypothesis on the final estimates, from the first to the third step. While these figures show that the estimation choices are rather conservative, they do not provide information on the overall impact of each of them. This is why I replicated the entire procedure, by taking at each step either the most conservative or the least conservative choice in order to see how they cumulatively affect my results. Figure 1.11 shows the decision tree of this procedure. Two facts stand out. First, there is a large uncertainty in the estimation procedure, with a difference of 10 percentage point between the lower bound and the upper bound estimate, which is not surprising given the data limitation. Second, despite this high uncertainty, the main conclusions of the paper remain unchanged, with lower bound estimates that show extreme levels of inequality (on average, the top 10 percent

32. Given the uncertainty surrounding the use of billionaires data, I only present averaged statistics over the period as the trends may not be reliable. In any case, the wealth share stay extremely high throughout the period, with a minimum for of 35 percent and 67 percent for the top 1 and 10 percent of the adult population (see Appendix A).

33. See Assouad et al. (2018). Derenoncourt (2018) interestingly say that a lack of transparency and data is itself a form of data, as the production of tax records depends on state capacity, fiscal infrastructures and therefore on the actual level of inequality in a society.

and 1 percent richest receive 50 percent and 19 percent of total national income).³⁴

1.3.6 Discussion : What are the drivers of such extreme levels of inequality?

How can we explain such extreme levels of inequality ? The results presented in this paper are consistent with scholarships in economics, political science and history that have highlighted various mechanisms amplifying inequality in Lebanon or in the region as a whole. This section briefly describes some features of the Lebanese political economy studied in the literature which can explain the extreme levels of income and wealth inequality we observe in the country.

The roots of such high levels of inequality can first be found in the Taif Agreement of 1989, which sealed the end Lebanese Civil War and reinforced the Lebanese "consociational democracy". This political arrangement, in which each sect is represented in national institutions and can influence policy seems attractive at first: Its announced goal is to guarantee that power is equitably shared among sects in order to prevent civil conflict (Lijphart, 1969). Research has shown however that such a political arrangement is has various perverse effects and is associated with bad governance outcome (Miguel, 2004; Spears, 2002). In Lebanon, the consociational democracy has enabled the formation of a "Party Cartel", a coalition of elites and parties that are ideologically opposed, but who have to share power. Such governments do not manage to implement cohesive and consistent public policies (Parreira, 2020). Another adverse outcome of a party cartel is that, despite their ideological opposition, party leaders manage to collude and prevent political competition. As a consequence, it is difficult for new political actors and parties to emerge, offer citizens alternatives, and make the cartel accountable. This has allowed the ruling elite to pass laws and policies favoring their economic and political privileges. The country has for example consistently opted for laissez-faire economic

34. I should emphasize that the upper and lower bound estimates are not realistic. For an example, the lower bound estimates assume that there is only 10 percent of capital income missing, while government reports on tax revenues suggest that they should be at least as high as 15 percent.

policies, resulting in the absence of welfare state and large-scaled redistributive policies but also major tax breaks for the wealthiest in following decades (Gaspard, 2004).³⁵. This is striking when we look at the tax system and the PIT in particular. Top marginal tax rates in Lebanon are quite low by international standards: Lebanon imposes its top earners much more lightly than France, Germany, the United Kingdom, and the United States, as well as countries from the Global South from 1979 to today (see Figure 1.12 Panel (A) and (B)). Similar conclusions can be reached for other taxes on capital income. This system also allowed the ruling elites to create mutually advantageous relationships with business elites. For instance, as many as eighteen of the country's twenty top banks have major shareholders linked to political elites, and that 43 percent of bank assets are subject to political control (Chaaban, 2016). This is significant because, historically, economic sectors dominated by a concentration of politically connected firms proved less competitive (Diwan and Haidar, 2021). As a consequence, a limited group of families has been able to share between themselves most sectors and enterprises of the country and extract large rents on virtually all economic activity at the expense of the majority, which makes the results on the distribution of income and wealth not surprising.

35. This however is not new: Lebanon has the oldest liberal market system in the region and its governance is characterized by minimal state interventions (Jawad, 2009). The only attempt to build strong public institutions and to create a welfare state occurred during Fouad Chehab's presidency between 1958 and 1964, contrasting with the liberal tendency prevailing since the independence, and which prevailed again afterwards. Since then, social welfare and state reforms have constantly been relegated to the background, while the Hezbollah ensures basic solidarity and redistributive functions and became a large clientelistic network if not a state within the state (Cammatt, 2014; Daher, 2014)

1.4 Conclusion

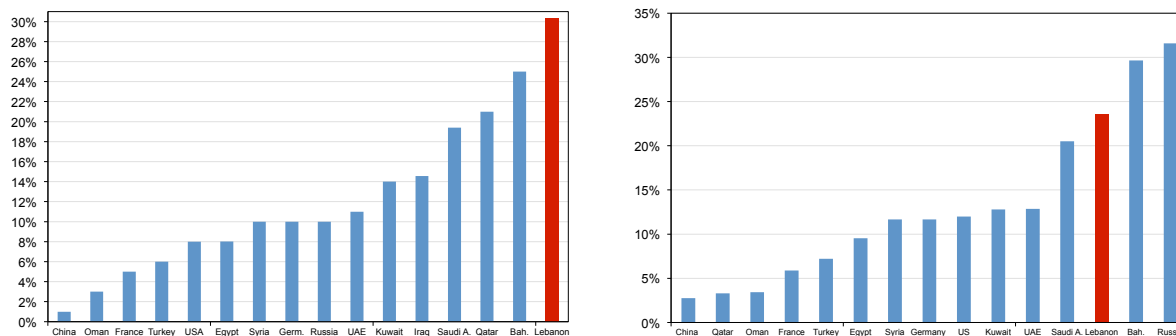
In this paper, I combine national accounts, survey, fiscal data and wealth ranking to estimate the national income distribution in Lebanon between 2005-2014. To the best of my knowledge, this paper is the first to use personal income tax records to study income inequality in a Middle Eastern country. I find that income and wealth are extremely concentrated and that the richest Lebanese caught the bulk of the national income growth under the period of study. These results put in perspective the so-called Lebanese economic miracle.

The main contribution of this study is to review available data sources on income and wealth and to combine them in a transparent manner to produce novel estimates of income inequality in Lebanon. To my knowledge, this paper provides the first reliable inequality statistics in the Middle East, to the extent that it uses administrative fiscal data to correct survey based estimates. It therefore contributes to answering the so-called "Arab Inequality Puzzle" and to shed light on the unequal concentration of economic resources that might be otherwise hidden by relatively high per capita average income. This study is however limited given the data at hand. In particular, it is at this stage difficult to reach robust conclusion on the dynamics of inequality and their roots. Yet, despite these data shortcomings, as there often exists in developing countries, one can still reach credible conclusions on income concentration, by indicating all the assumptions made and their effects on the final series.

Figure 1.1 – Average Billionaires’ Wealth as Share of National Income in Selected countries

(a) 1990-2016

(b) 2005-2016



Notes: Total billionaire wealth as a share of total national income (measured at market exchange rates), average over for 1990-2016 (a) and for 2005-2016 (b). For 1990-2005 Lebanon is ranked second below Qatar, with an average of 33 %. Author’s computation using rich lists from Forbes and Arabian Business magazines, for Middle Eastern countries.

Figure 1.2 – From Survey to Taxable and Total National Income, 2005-2014

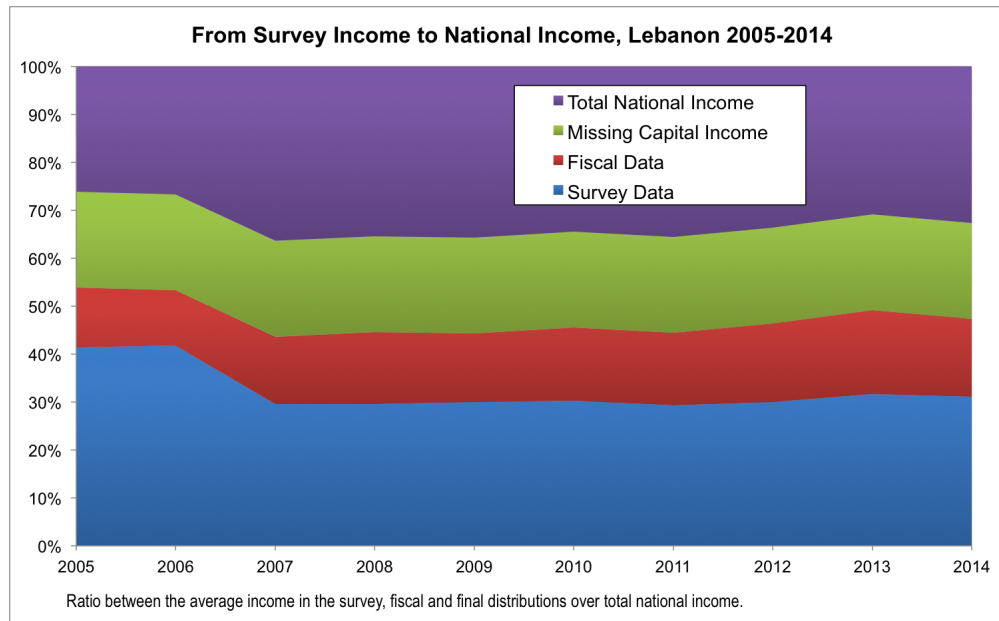
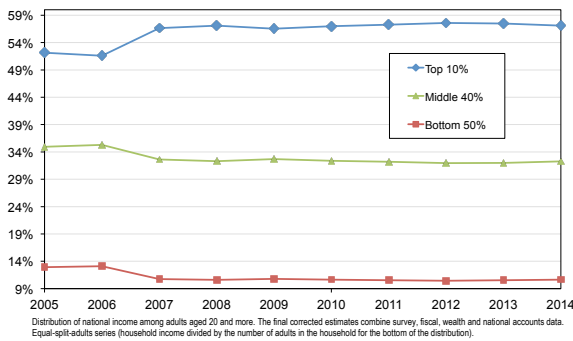
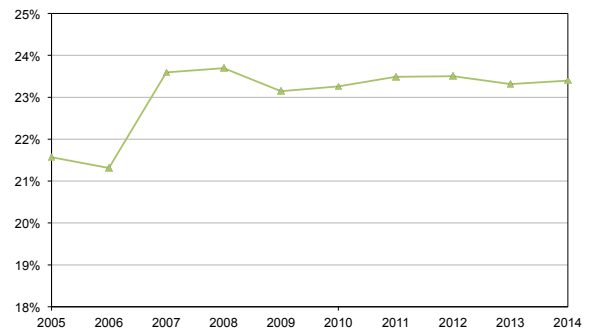


Figure 1.3 – Income Shares in Lebanon, 2005-2014

(a) Top 10%, Middle 40% and Bottom 50% income shares



(b) Top 1% income share



Notes: Distribution of national income among adults aged 20 and more. The final corrected estimates combine survey, fiscal, wealth and national accounts data. Equal-split-adults series (household income divided by the number of adults in the household for the bottom of the distribution).

Figure 1.4 – Income Shares in Lebanon, 2005-2014: Top 0.1% vs. Bottom 50%

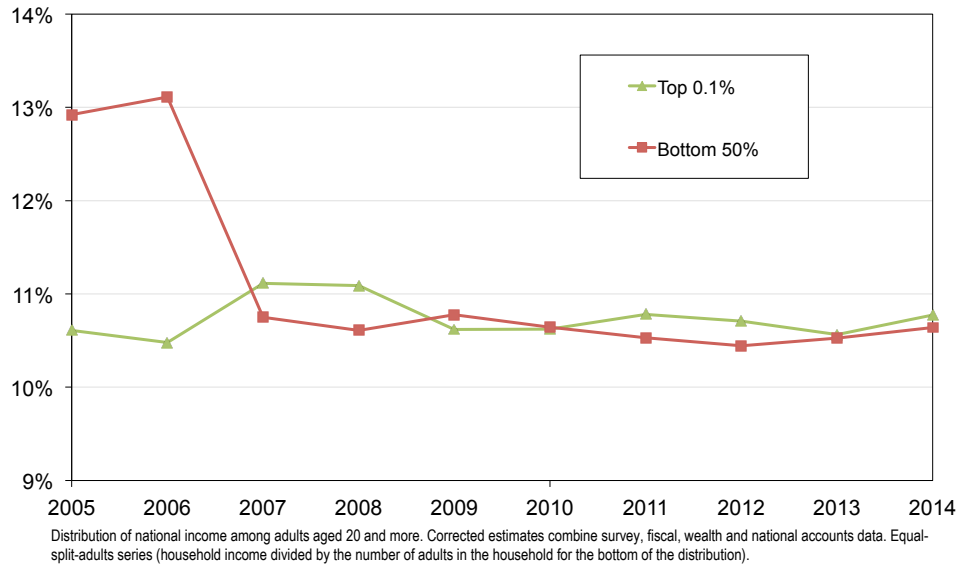
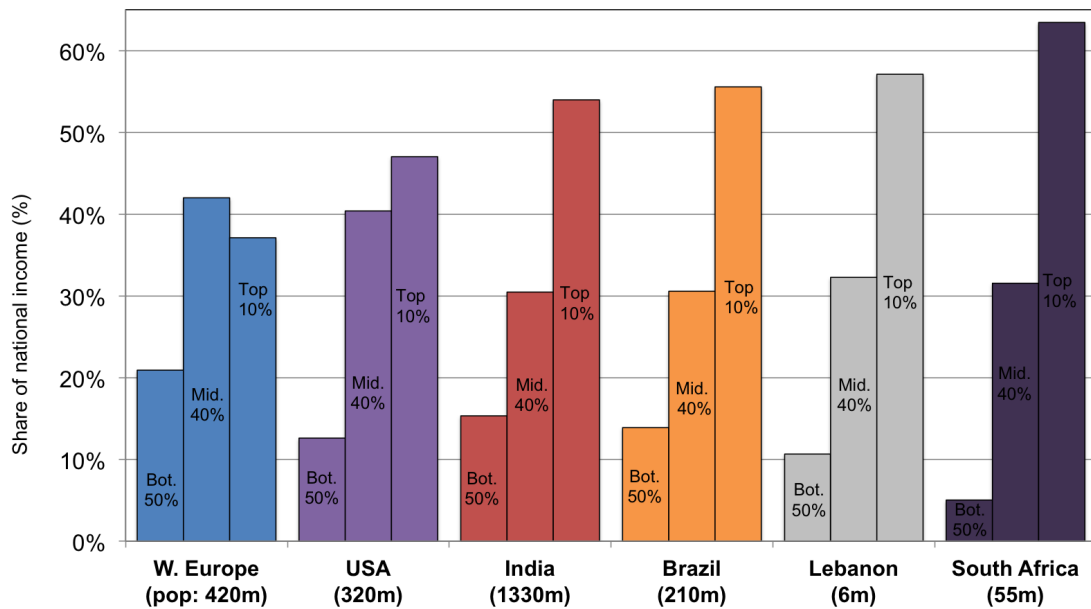


Figure 1.5 – Income Shares in Lebanon, 2005-2014: Top 0.1% vs. Bottom 50%



Source: [World Inequality Database](#). For Lebanon: Distribution of national income among adults aged 20 and more. The final corrected estimates combine survey, fiscal, wealth and national accounts data. Equal-split-adults series (household income divided by the number of adults in the household for the bottom of the distribution).

Figure 1.6 – Population vs. income cumulative growth since 2005

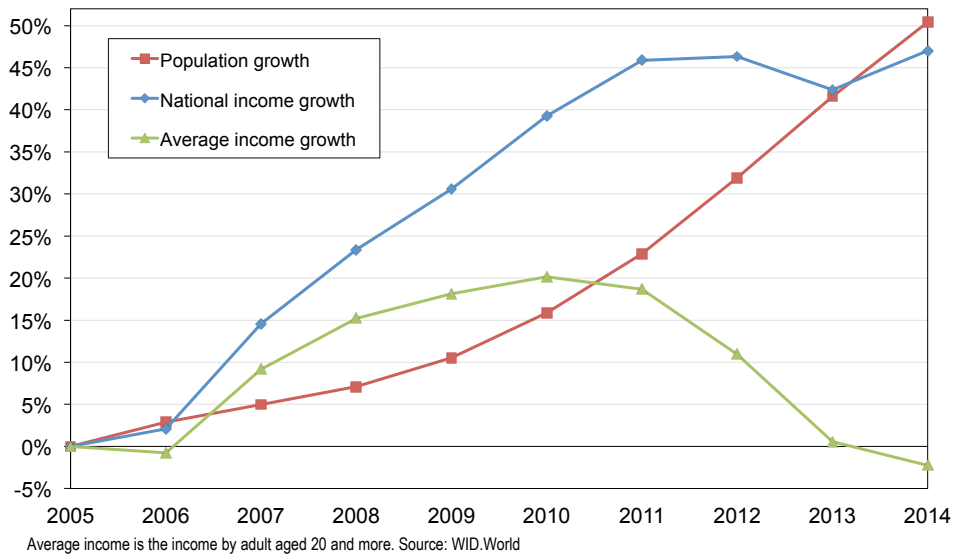


Figure 1.7 – Cumulative real growth by percentile, Lebanon 2005-2014

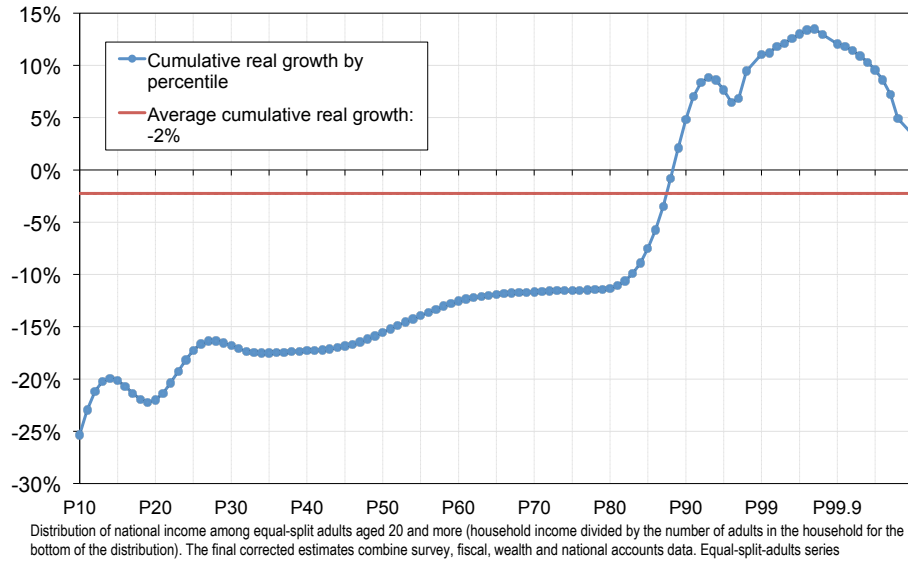
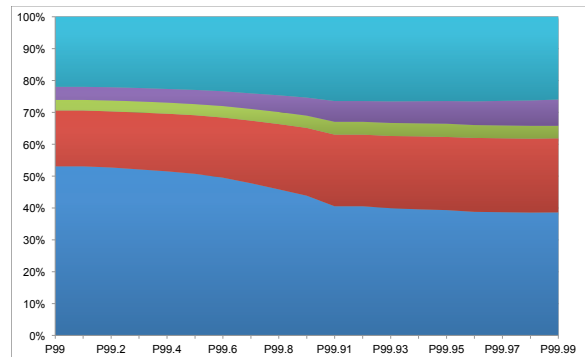
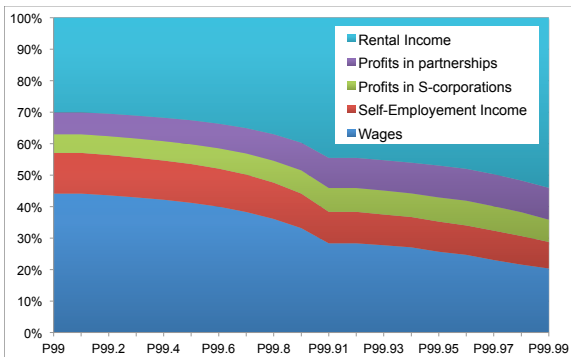


Figure 1.8 – Composition of top income by income categories: 2005, 2014

(a) 2005

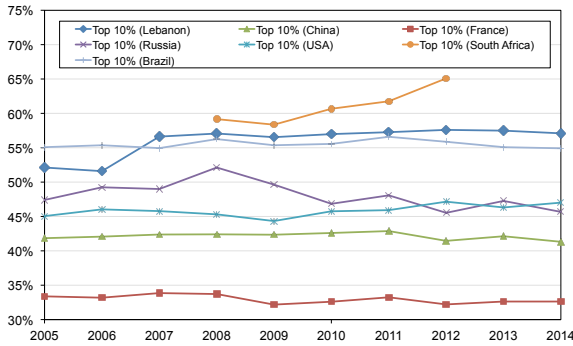
(b) 2014



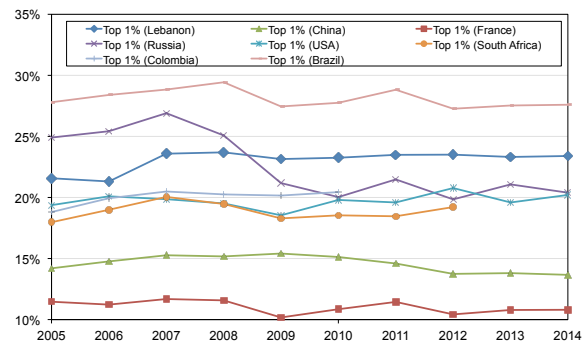
Source: Author's computation using the fiscal micro files.

Figure 1.9 – Top income shares: Lebanon vs. Selected countries, 2005-2014

(a) Top 10% income share



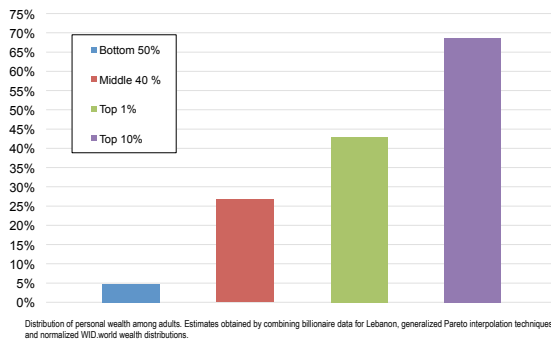
(b) Top 1% income share



Notes: Distribution of pretax national income (before taxes and transfers, except pensions and unempl. insurance) among equal-split adults (income of married couples divided by two) for all countries except South Africa. For South Africa, distribution of fiscal income. Sources for Brazil, China, Colombia, France, Russia, South Africa and USA: WID.world.

Figure 1.10 – Wealth inequality in Lebanon and in selected countries

(a) Lebanon: average over 1990-2016



(b) Lebanon vs. Selected countries: average over 2005-2014

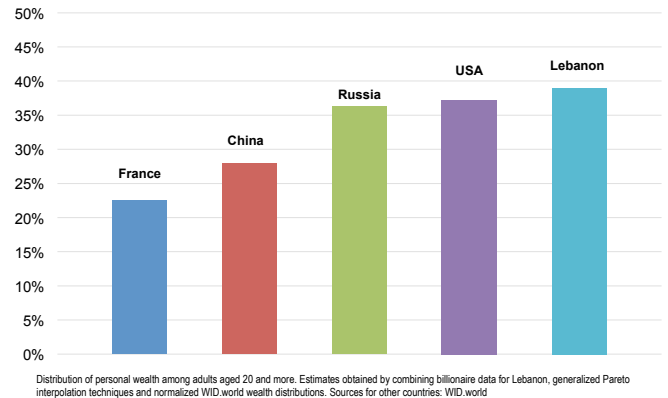
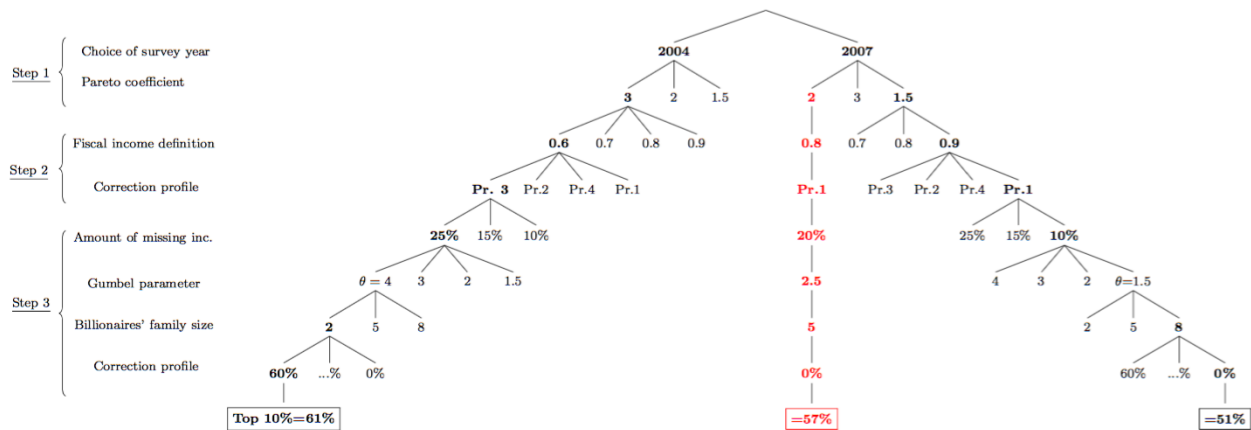
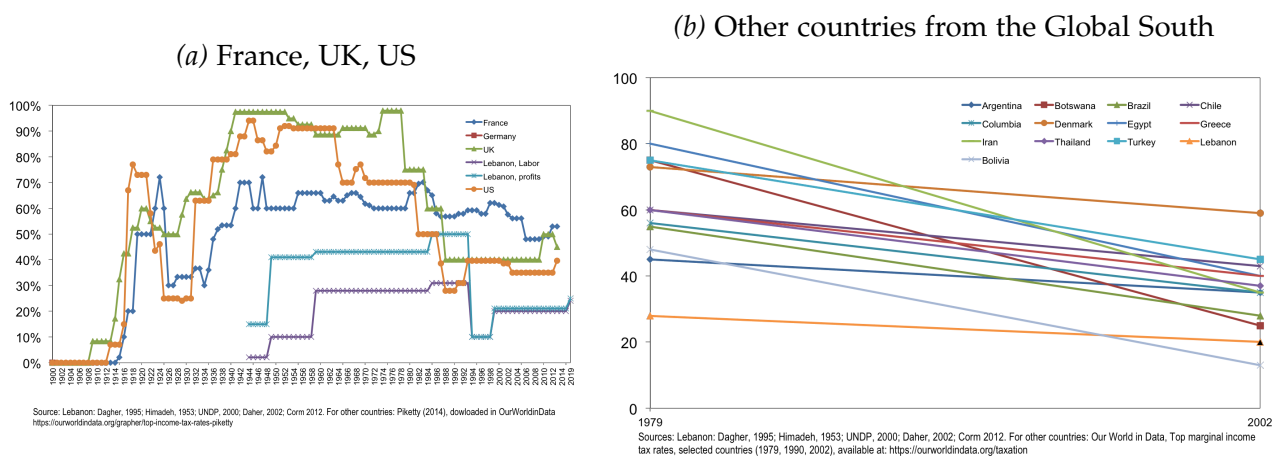


Figure 1.11 – Decision tree of the estimation procedure and implications for inequality, average for the 2005-2014 period



Source: Author’s computations. The decision tree explicits the main hypotheses made in the estimation procedure: (1) Choice of the survey year; (2) Choice of the Pareto coefficient in the last income bracket of the survey tabulation to estimate the survey distribution with gpinter; (3) Fiscal income definition ($taxableincome = 0.8 * fiscalincome$ etc.); (4) Correction profile used to link the survey data (at the bottom) to the fiscal data (at the top of the distribution); (5) Amount of missing capital income to reallocate; (6) Gumbel parameter, that defines the correlation structure between fiscal and non-fiscal income; (7) Billionaires’ family size; (8) The weight of billionaires’ wealth in the distribution: 20% means 20% of the correction factor adjustment is achieved at P99.9 etc. See appendix A.1 for details on each hypothesis. At each node, the options on the left corresponds to the most unequal hypothesis. The branch on the left corresponds to the upper bond estimate, on the right to the lower bound estimate. The assumptions in red define my benchmark specification.

Figure 1.12 – Evolution of Top marginal tax rates in Lebanon versus Selected countries



Sources: Lebanon: Dagher, 1995; Himadeh, 1953; UNDP, 2000; Daher, 2002; Corm 2012. For other countries in Panel A: Piketty (2014), downloaded in OurWorldinData. For countries in Panel B: “Top marginal income tax rates, selected countries” (1979, 1990, 2002), downloaded in OurWorldinData.

Table 1.1 – Income thresholds and income shares in Lebanon, 2016

Income groups	Number of adults	Income thresholds	Average income	Income share
Full population	3,717,891	0	14,356	100.0%
Bottom 50%	1,858,946	0	3,055	10.6%
Middle 40%	1,487,156	5,977	11,577	32.3%
Top 10%	371,789	29,373	81,978	57.1%
<i>incl. Top 1%</i>	37,179	123,651	335,930	23.4%
<i>incl. Top 0.1%</i>	3,718	453,700	1,593,622	11.1%
<i>incl. Top 0.01%</i>	372	2,224,880	8,593,634	6.0%
<i>incl. Top 0.001%</i>	37	11,782,820	47,365,937	3.3%

Notes: Statistics on the distribution of income expressed in PPP Euro 2016. Adult individual aged 20 and more; Equal-split assumption among adult members of a household. In 2016, 1 euro = 1641 LBP (market exchange rate) or 172.7 pound (PPP). Income corresponds to pre-tax national income. Fractiles are defined relative to the total number of adult individuals in the population. Corrected estimates (combining survey, fiscal, wealth and national accounts data), from 2014 adjusted for the price change between 2014-2016 (shares are not affected).

Table 1.2 – Average incomes in Western Europe, USA, Brazil, India and South Africa: 2016 Euros (PPP)

Income groups	USA	Western Europe	Brazil	South Africa	India
Full population	61,795	34,214	9,115	8,439	4,391
Bottom 50%	15,572	14,308	2,233	848	1,345
Middle 40 %	62,387	35,916	7,387	6,654	3,343
Top 10%	290,542	126,938	50,432	53,538	23,808
<i>incl. Top 1%</i>	1,248,259	417,501	253,759	154,877	95,388
<i>incl. Top 0.1%</i>	5,759,294	1,553,248	1,313,729	486,861	378,319
<i>incl. Top 0.01%</i>	26,899,363	6,143,396	6,817,909	1,457,794	1,684,895
<i>incl. Top 0.001%</i>	117,410,496	24,494,358	35,399,859	4,286,839	17,278,335

Notes: Statistics on the distribution of income expressed in PPP Euro 2016. Adult individual aged 20 and more; income of married couples is split into two. Income corresponds to pre-tax national income. Fractiles are defined relative to the total number of adult individuals in the population. Corrected estimates (combining survey, fiscal, wealth and national accounts data).

Source: Assouad et al. (2018)

2. MEASURING INEQUALITY IN THE MIDDLE EAST 1990-2016: THE WORLD'S MOST UNEQUAL REGION?

1

1. This is joint work with Facundo Alvaredo and Thomas Piketty

2.1 Introduction

This paper lists and combines all available data sources on income and wealth in the region (national accounts, household surveys, income tax data, and wealth rankings) in order to provide novel estimates of the distribution of income in the Middle East between 1990 and 2016. According to our benchmark series, the Middle East appears to be the most unequal region in the world, with a top decile income share as high as 61%, as compared to 36% in Western Europe, 47% in the USA and 55% in Brazil. This is due both to enormous inequality between countries (particularly between oil-rich and population-rich countries) and to very large inequality within countries (which we probably under-estimate, given the limited access to fiscal data).

These estimates are based upon two methodological innovations. To our knowledge, our paper is the first attempt to combine Middle East household surveys with income tax data. Namely, we use the findings from Lebanese income tax micro-files recently exploited by [Assouad \(2021\)](#) and apply generalized Pareto interpolation techniques ([Blanchet et al., 2022](#)). This leads us to significantly correct upward standard survey-based, within-country inequality estimates. This upward correction should still be viewed as a lower bound, but more plausible than usual measures based solely upon self-reported data. Next, our paper is also the first attempt to combine within-country inequality measures in order to estimate the distribution of income for the entire Middle East region. Both innovations play an important role in accounting for our high inequality findings (they both have impacts that are comparable in magnitude). We stress that we still face important limitations and uncertainties regarding the measurement of income distribution in the region, and that increased transparency on income and wealth is highly needed. However our main conclusion - namely the fact that the Middle East is one of the most unequal regions in the world, if not the most unequal region - appears to be robust.

Of course, we do not pretend that this high inequality level is the only explanation for the regional political instability. Many other factors - religious, cultural and political - certainly play an important role as well. But we believe that inequality can be part of the explanation, or at least that it belongs to a set of background factors that can contribute to generate political upheavals. The 1990 invasion of Kuwait by Iraq - two countries with vastly different per capita income and wealth - is a clear and extreme example. More generally, one can plausibly argue that perceptions about inequality and the fairness or unfairness of the distribution of income are determined not only by within-country inequality but also by inequality at the regional level (or sometime at the global level). To a large extent, this paper can be viewed as an exercise of aggregation. We show how changing the geographical level of analysis affects the measurement of inequality. In the case of the Middle East, the concept of nation-state may not be the most meaningful lens through which we can analyze the concentration of income. The total population of the region (about 410 million in 2016) is comparable to Western Europe (420 million) or the United States (320 million), and is characterized by a relatively large degree of cultural, linguistic and religious homogeneity (at least as compared to these other world regions). Therefore we feel that such comparisons are legitimate and to some extent informative - at least as much as the usual inequality comparisons between nation-states. Both types of comparisons seem to capture complementary and valuable dimensions of individual perceptions.

This paper is part of a broader project, namely the World Wealth and Income Database (WID.world), that attempts to produce annual distributional statistics - and possibly micro data on income and wealth distributions - that are comparable across countries ([Alvaredo et al., 2016](#)). For this, we follow a common methodology that involves the combination of national accounts, surveys, and fiscal data in a consistent manner to produce distributional national accounts. The methodology was already applied for the United States ([Piketty et al., 2016](#); [Saez and Zucman, 2016](#)), France ([Garbinti et al., 2018](#)),

China (Piketty et al., 2017) and Russia (Novokmet et al., 2017). Although there are similarities across countries regarding methods, lessons can be drawn from country-specific cases to help produce new databases for future works, in a context of scarcity of data.

The rest of this paper is organized as follows. In section 2, we relate our work to the existing literature on income inequality in the Middle East and at the global level. Section 3 describes our main data sources, concepts, and methodology. In Section 4, we present our main results on the evolution of income inequality in the Middle East, and we also compare our series to other countries. Section 5 provides concluding comments. This paper is supplemented by an extensive online appendix that includes all our raw data and codes and also presents additional results and robustness checks.

2.2 Literature Review: Inequality Measurement in Middle East countries

The study of the evolution of income and consumption inequality using household surveys is a well-established tradition in a number of countries in the Middle East (Said, 2007; Wahba, 1996). In addition, following the Arab Spring movement, there has been renewed interest in inequality measurement in Middle East countries. A number of papers have argued that income inequalities within these countries do not seem to be particularly high by international standards, and therefore that the source of dissatisfaction might lie elsewhere (Bibi and Nabli, 2009; Hlasny and Verme, 2018). This somewhat surprising fact, coined “the Enigma of Inequality” (UNDP, 2002) or the “Arab Inequality Puzzle” (World Bank, 2016), has produced a rising literature on inequality in the region (Assaad et al., 2018; Hassine, 2015; Hlasny and Verme, 2018; Ncube and Anyanwu, 2012). As noted in the introduction, our contribution to this literature is twofold. We combine household surveys with income tax data in order to correct upwards the top of survey-based income distributions, and we aggregate within-country distributional data in order to estimate the distribution of income at the level of the entire Middle East. This leads us to relatively novel (though not entirely unexpected) conclusions regarding extreme inequality in the Middle East.

We stress that these results should be viewed as exploratory and suffer from many limitations. In particular, despite our best efforts, our ability to properly measure income inequality within individual countries is severely limited by the low quality of available data sources. The problem is particularly acute in the Gulf countries, for which there exist very few studies on income distribution (El-Katiri et al., 2011), and where the low official Gini coefficients reported seem to contradict important aspects of their political economy, namely the growing share of migrant population, a large majority of which is composed by low-paid workers living in difficult conditions (Human Rights Watch, 2013). The flow of migrant workers in Gulf countries has grown

substantially over the period. This sharp increase contributed to make nationals willing to defend their numerous privileges, beginning by restraining naturalization.² But the most striking manifestation of the restrictions imposed to the migrant population is probably the highly exploitative “sponsorship system” of labor or “kafala system” [Kapiszewski \(2006\)](#), resulting in the creation of an extremely polarized social structure with two different groups in the legal, social and economic dimensions ([Kinninmont, 2015](#)). As far as we know, little research has been conducted to study the two populations in order to measure income inequality within Gulf societies. In the context of this paper, we attempt to put together all existing statistical information that has been published regarding the inequality of income between nationals and foreign workers in Gulf countries (see section 4 below). Unfortunately, we still face important limitations in our empirical and quantitative understanding of these issues.

Finally, our paper is closely related to the literature on the world distribution of income ([Bourguignon and Morrisson, 2002](#); [Lakner and Milanovic, 2013](#); [Milanovic, 2002](#)). In particular, [Lakner and Milanovic \(2013\)](#) attempts to correct upwards the top income share estimates constructed on the basis of national household surveys to study how much this impacts the measurement of the world distribution of income. Our approach is similar, except that we focus on regional inequality (i.e. inequality measured at the level of a broad region such as the Middle East) rather than global inequality. Of course both approaches are highly complementary: before we can perform a meaningful aggregation at the world level, it is important to ensure that we are able to do it at a broad regional level.

2. “Oil wealth, and the practice of dividing some of this among citizens, mean there are very strong economic incentives to limit citizenship to a small pool of people. Gulf nationals typically do not pay income tax, have free health care and education provided by the state, receive subsidies for electricity and fuel, and often receive other benefits (such as land grants). Traditionally they have also expected the state to provide a job - an idea enshrined in some Gulf constitutions and - housing.”([Kinninmont, 2015](#))

2.3 *Data sources, Concepts and Methodology*

This paper relies on four types of data sources: household surveys, income tax data, wealth rankings and national accounts. We define the Middle East as the region going from Egypt to Iran, and from Turkey to the Gulf countries. We start by putting together a macroeconomic database including annual series on population and national income between 1990 and 2016. Basic descriptive statistics for 2016 are reported in Table 2.1. The region is characterized by very large between-country inequality (we further discuss this issue in section 4). All details about the data sources and methods used to construct homogenous national accounts are described in the online appendix.

In order to estimate the distribution of income in the Middle East, our general methodology follows three steps. We begin with the Middle East household income surveys data series (step 1), which we correct using (i) generalized Pareto interpolation techniques (Blanchet et al., 2022) and (ii) personal income tax micro-data available for Lebanon (step 2) (see Assouad (2021) for a description of these fiscal data and their main limitations) We then use national accounts and rich lists in order to impute tax-exempt capital income (step 3). Our concepts and methods generally follow those described in the Distributional National Accounts guidelines used for the World Wealth and Income Database (Alvaredo et al., 2016). In particular, the methodology in three steps is very similar to that used for China in Piketty et al. (2017), and for Russia in Novokmet et al. (2017), with some differences highlighted in the following sections.

2.3.1 *First step: constructing a household income database for the Middle East*

Income and inequality data are scarce in the Middle East, notably in the poorest and the richest countries. Although many national statistics offices undertake household surveys on income or expenditure, access to the data is very limited. Until recently, it

was almost impossible to obtain micro-data.³ Finally, when they exist, the databases are often of poor quality (see [Bibi and Nabli \(2009\)](#) for a review of existing data, and an assessment of their access and quality).

The first part of our work consists in gathering available sources to create a Middle East income database and generate raw survey-based inequality series at the national and then the regional levels. [Table 2.2](#) summarizes the years on which household survey data were available: there is for each country between 1 and 16 years with data. Regarding the format, there are four cases: (1) 8 countries with survey micro-data; (2) 6 countries with tabulated information on the distribution of income, extracted from household surveys reports and/or statistics offices publications; (3) 3 countries with data on expenditure and consumption (tables or micro-data); (4) 1 country, Saudi Arabia, with no detailed published data. In the online appendix we provide a thorough description of all data sources country by country, the information available, and the methodology used to combine them to produce income distribution series over the 1990-2016 period ([Appendix A](#)). We briefly summarize three main issues regarding the data construction process in the following.

A first issue concerns the definition of income. Unfortunately, the data quality makes it impossible to harmonize the series in a completely satisfactory manner. Only the micro-data for Turkey contain relatively detailed information on income categories (wages, pension and other replacement income, business, and capital income) enabling to distinguish between different income concepts. Other micro-databases only provide total disposable income, with however some additional information on imputed rental income and/or the amount of taxes on consumption and durable goods, property taxes etc. paid for some years and countries. Tabulated data usually contain limited information on the definition of income. Whenever possible, the survey income concept that we use attempts to approach pre-tax, post-replacement income (see [DINA guidelines](#),

3. See in particular the “Open Access Micro Data Initiative” undertaken by the Economic Research Forum.

[Alvaredo et al. \(2016\)](#)). More precisely, pension income (and other replacement income such as unemployment insurance) is included, while pension contributions (and other social contributions financing replacement income flows) are deducted. Therefore, in the trade-off between harmonizing our database (between years and/or countries) and approaching the pre-tax income concept we choose the latter. This is a substantial limitation that needs to be corrected in the future.

The second issue concerns the unit of observation. We take the adult individual as the basic unit, and we assume that income is equally split between adult household members ([Alvaredo et al., 2016](#)). We are therefore interested in the distribution of per-adult equal-split income.⁴ We normalize our series to the adult population (i.e. aged 20 and more). Using the generalized Pareto interpolation techniques developed in [Blanchet et al. \(2022\)](#) and the `gpinter` web interface (www.wid.world/gpinter), we estimate the full distribution of raw survey income separately for all countries and for the region as a whole.⁵ We express the distributions in terms of generalized percentiles (or *g*-percentiles).⁶

The third issue is related to the years without data. As one can see from [Table 2.2](#), household surveys are available only for a limited number of years.⁷ To infer the distribution of years with no data, we use the household surveys distribution of the closest available years. For a number of countries, we only have one household survey, which means that by construction we are forced to use the same inequality level over the entire 1990-2016 period. As we repeatedly stress throughout the paper, this major limitation implies that we cannot draw robust conclusions about the evolution of income

4. See Appendix A for more details on the country specific hypothesis made to derive the per adult income distributions.

5. We use the merging option to derive the national distribution of Iran (merging rural and urban distribution) and of the Gulf countries (merging the foreigners and non-foreigners distribution).

6. There are 127 *g*-percentiles: 99 for the bottom 99 percentiles, 9 for the bottom 9 tenth-of-percentiles of the top percentile, 9 for the bottom 9 one-hundredth-of-percentiles of the top tenth-of-percentile, and 10 for the 10 one-thousandth-of-percentile of the top one-hundredth-of-percentile.

7. We also constructed estimates based on the assumption of linear inequality trends between survey years. This made very little difference in both the level and trend obtained for total Middle East inequality, so in our benchmark series we simply use the closest available year for country-level data.

inequality: the main objective of this paper is to estimate the overall level of income inequality in the Middle East, not the evolution.

Finally, and maybe most importantly, in order to ensure maximal comparability across countries and time, we choose to anchor all country-year-level income distributions to the relevant per adult national income. That is, for every country-year pair, we proportionally upgrade all income levels for all percentiles so that per adult average income coincides always with per adult average national income observed in our macroeconomic database (therefore keeping the income distribution and shares constant). By doing so, we certainly do not pretend that available national income series are perfectly comparable. We simply assume that these are the most comparable income series we have: national accounts at least attempt to apply the same definition of national income in all countries (as defined by the SNA Guidelines developed under the auspices of the UN and other international organizations), which is not the case with survey income. This issue is further discussed in the DINA Guidelines ([Alvaredo et al., 2016](#)).

We also report on [2.2](#) the ratios between total survey income and the national income for the different countries. For most Middle East countries, aggregate ratios are around 40%-50%, which is fairly small, but not unheard of by international standards. Note however that the ratios are substantially smaller in Gulf countries - as low as 20%-30%. That is, compared to other countries, a very large fraction of national income of Gulf countries is missing from self-reported household survey income. To the extent that missing income components benefit relatively small groups of the population, this implies that we are likely to severely underestimate income inequality within Gulf countries (see section 4 below for a discussion).

2.3.2 Second step: Fiscal data correction

Self-reported survey data is well-known to underestimate incomes at the top (say, within the top decile, and particularly within the top percentile). Generally speaking,

the strategy followed in the World Wealth and Income Database (WID.world) in order to correct for this is to use income tax micro-files (together with national accounts and wealth data in order to cover tax-exempt income). In case income tax data do not exist or is limited in scope (e.g. in case one can only access income tax tabulations rather than micro-files), the DINA Guidelines recommend to supplement existing data with generalized Pareto interpolation techniques ([Alvaredo et al., 2016](#); [Blanchet et al., 2022](#)).

In the case of Middle East countries, income tax data are unfortunately extremely limited. Lebanon is the only country for which we were able to access income tax micro-files. These data are relatively detailed, consisting on yearly quasi-exhaustive micro-files over the 2005-2014 period (see [Assouad \(2021\)](#) for a detailed description of this data source). However, for other countries, despite our best efforts, we do not have any income tax data (not even income tax tabulations).

This is unfortunate, because household surveys in the Middle East appear to underestimate top incomes at least as much as in the rest of the world, and possibly more. In particular, survey-based inverted Pareto coefficient $b(p)$ are implausibly low for top incomes, generally around 1.5-1.7 (and sometime even less than 1.5) at the level of the top 10% (i.e. $p=0.9$).⁸ In contrast, in all countries in the world with reliable income tax data, inverted Pareto coefficients $b(p)$ are typically between 2 and 3 (or even more in high inequality countries), and tend to follow a U-shaped generalized Pareto curve, with a rising part within the top decile. The Lebanese income tax micro-files confirm this general finding: top income levels reported in tax data are much higher than in household surveys (top 1% incomes are typically 2-3 times higher, with large variations across income levels and over years), and the tax-corrected inverted Pareto coefficients within the top decile are around 3 or higher ([Assouad, 2021](#)). The reasons why household surveys almost systematically lead to excessively low b coefficients typically come from the fact that surveys suffer from various under-reporting, truncations and top cod-

8. See on-line technical appendix, Table A3.

ing problems (with top coding, or self-censored top incomes, b naturally becomes very close to 1 at the very top).⁹ Naturally, surveys have other merits, and include detailed socio-demographic information that one could never obtain using tax data. However, for the study of the top decile - and also for the study of the total inequality level of a country, given the importance of the income share going to the top decile - it is necessary to supplement surveys with other sources and methods.

In order to construct our benchmark series, we choose to adopt correction factors that are based upon the income tax data from Lebanon. More precisely, the income tax micro-files enable us to compute correction coefficients for thresholds and upper average income by g -percentiles.¹⁰ We apply no correction below $p=0.8$, i.e. we assume correction factors exactly equal to 1 below the top 20%, which is approximately the case in the Lebanese data (see [Assouad \(2021\)](#) on the choice of profiles). These coefficients do not depend on the Lebanese income level, but only on the percentile. We apply the average correction coefficient per percentile over the 2005-2014 period in Lebanon to all other countries. We have also computed a large number of variant series based upon alternative assumptions (see on-line appendix). The impact on the overall inequality level in the Middle East and the comparison with other world regions is relatively limited (as a first approximation). In order to derive more precise estimates, we would need to have access to income tax data (at least in the form of income tax tabulations, and ideally in the form of micro files) for all Middle East countries.

9. [Hlasny and Verme \(2018\)](#) Figure 10, p.28) use household income surveys for Egypt between 1999 and 2010 and argue that top-decile inverted Pareto coefficients around 1.5-1.7 are not unusual by international standards. However this conclusion comes entirely from the fact that they compare with coefficients coming from household surveys (which are artificially low).

10. In particular, to derive the raw survey distribution of Lebanon, one needs to assume an inverted Pareto coefficient at the top due to the format of available tabulations. This affects the correction coefficients and the levels of inequality in all countries. For other variants on the definition of income and profile of correction. We have also computed variant series based on the assumption that inverted Pareto coefficients $b(p)$ take average WID.world values (typically within the interval [2,3]) for Middle East countries other than Lebanon. This leads to results for total Middle East inequality that are close in magnitude to those presented here (see on-line technical appendix, Tables A3-A4 for inverted Pareto coefficients for the various countries and years, before and after our benchmark fiscal corrections).

2.3.3 *Third step: Missing capital income and wealth correction*

Finally, we correct our fiscal income series to take into account non-reported and tax-exempt capital income. Important components of capital income are missing from fiscal income data, even in the absence of any tax evasion [Alvaredo et al. \(2016\)](#); [Piketty et al. \(2017\)](#). They typically include corporate retained earnings and imputed housing rental income. We assume these “non-fiscal” income y_{nf} is equal to 10% of national income in each country, a reasonable figure given our findings in other countries. For Lebanon however, we estimate it to be 20% of national income, by using available information from national accounts and government reports on tax revenues, published by the Ministry of Finance ([Assouad, 2021](#)). Then, to estimate the distribution of personal income $y_p = y_f + y_{nf}$, i.e. the sum of fiscal and non-fiscal income, we need to make an assumption about the distribution of “ y_{nf} ” and the correlation between y_f and y_{nf} . We assume that y_{nf} follows the same distribution as wealth, which we estimate by applying generalized Pareto interpolation techniques to household wealth surveys and wealth rankings (see below). As for the correlation structure between y_f and y_{nf} , on the basis of estimates obtained in countries with adequate micro-files, we use the family of Gumbel copulas, with Gumbel parameter $\theta = 3$ ([Novokmet et al., 2017](#); [Piketty et al., 2017](#)). We should stress that this wealth-based correction has a relatively limited impact on our final income inequality estimates (and in particular a much smaller impact than the fiscal data correction), so that the uncertainty that we are facing here is unimportant for our main findings (see section 4, Figures [2.18](#)- [2.20](#)).

In order to estimate wealth inequality, we proceed as follows. Most observers tend to assume that the level of wealth inequality in the region is high by both international and historical standards. However, there is substantial uncertainty about the exact level of wealth concentration, due to the almost complete lack of proper statistical evidence. Here we follow a simple methodology similar to that applied in ([Novokmet et al., 2017](#)) and use rich lists to produce estimates of top wealth shares for Middle Eastern countries

in 2016, which we then use to allocate tax-exempt capital income.

We use billionaire's lists published by Forbes and the magazine Arabian Business. Generally speaking, we find that the share of billionaires' wealth in national income is indeed extremely high by international standards. For Saudi Arabia, Qatar, Bahrain and Lebanon, it is greater than 20% on average, while total billionaire wealth represents between 5% and 15% of national income in the United States, Germany and France over 2005-2015. Wealth concentration is particularly high in Lebanon, where the average income and the average wealth are substantially below Western levels ([Assouad, 2021](#)). We stress however that billionaires' lists are particularly fragile and volatile in the Middle East. There are relatively few billionaires and their number varies substantially from year to year (many years have no data). For instance, Forbes reports one or two billionaires in Bahrain and Qatar and only in three years between 1990-2016. However, for some years, billionaires' wealth can represent a very high share of national income. Given that the figures are extremely volatile, using this data source to identify a trend in wealth concentration is impossible. Several reasons can explain why wealth rankings might be particularly incomplete in the region. First, large amounts of wealth may be missing due to a pervasive use of tax havens and offshore bank accounts. The data leaked from HSBC Switzerland and Mossack Fonseca (the so-called "Swiss leaks" and "Panama Papers") show that Middle East countries are among the top clients of those offshore financial institutions. Evidence indeed indicates that hidden wealth is high by international standards ([Zucman, 2013](#)). [Andersen et al. \(2017\)](#) also show that "petroleum-rich autocracies" in the Arab world tend to hide larger amounts of wealth and that they would do it more easily than other countries with oil resources. In addition, rich lists do not include wealth owned by ruling families and heads of states. This may lead to a substantial downward bias in the region, where the line between public and private property is often blurred. We attempt to include figures on state leaders'

wealth when we could find some, but reliable information is very scarce.¹¹

For all these reasons, we did not attempt to derive annual wealth distribution series. Rather, we compute one average estimate for wealth inequality for each country, applying the same general method as in [Novokmet et al. \(2017\)](#) for Russia. For each country, we compute average billionaire wealth (as a fraction of national income) over all years available in the 1990-2016 period. We then compute average standardized distributions of wealth for the US, France and China from WID.world series. We note that variations across countries and over time in these standardized wealth distributions mostly happen above $p_0 = 0.99$, i.e. below $p_0 = 0.99$ the ratios of the different percentile thresholds to average wealth are relatively stable over time and across countries, at least as a first approximation with most of the variation taking place within the top 1%. Therefore we choose to use the same normalized distribution for Middle East countries below $p_0 = 0.99$ as the average US-France-China normalized distribution. To estimate the average wealth, we compute an annual average wealth-income ratio over all countries available in WID.world, and we apply this average to each country average income. The difficult question is to know how to link the distribution from $p_0=0.99$ to billionaire level, and also to make an assumption about the average number n of adults per billionaire family (sometime Forbes includes very large family groups in the same billionaire family; sometime it is just one individual or one married couple). We first re-estimate the 127 generalized percentile within the top 1% of the normalized distribution in order to reach billionaires' level. In our benchmark series we assume $n = 5$ and a linear correction factor $f(p)$ from $p_0 = 0.99$ up to billionaire level, as this assumption seems to work relatively well for the US, France and China. This method gives a first approximation of the concentration of wealth in the region. In the appendix we present a number of alternative series based upon explicit assumptions and generalized Pareto interpolation

11. We gathered some figures from various sources (newspapers articles, Forbes' "Royals" and "Dictators" lists). We could not cover all ruling families and, when we find information, it is only available for some years. For an example, figures on the Assad family's wealth are only available for two years. We did not find figures on billionaires in Jordan.

techniques.

2.4 *Main Results: Extreme Concentration of Income in the Middle East*

We now present our main results on the level and evolution of income inequality in the Middle East. We start by describing the general evolution of average incomes and between-country inequality in the Middle East region over the 1990-2016 period. We then present what we consider our most robust and interesting finding, i.e. the extreme level of income concentration in the Middle East as a whole (as compared to other world regions), taking into account both between-country and within-country inequality. Finally, we discuss our findings regarding the evolution of income inequality in the Middle East over the 1990-2016 period (which, as we stressed in the previous section, should be viewed as more fragile and exploratory than our findings regarding the level). Complete series and detailed country-level estimates are available in the online appendix.

2.4.1 *Evolution of average incomes and population in the Middle East*

The 1990-2016 period has seen a rapid population growth in the Middle East: total population rose by about 70%, from less than 240 million in 1990 to almost 410 million in 2016. The rise in average income has been much more modest. Using purchasing power parity estimates (expressed in 2016 euros), per adult national income rose from about 20 000 Euro in 1990 to 23 000 in 2016, i.e. by about 15%. Using market exchange rates (again in 2016 euros), per adult national income rose from less than 9000 Euro in 1990 to about 10000 Euro in 2016 (see Figure 2.1).

Given the importance of migrations and economic relations between the two regions, it is natural to compute the ratio between per adult national income in the Middle East and the West European average (itself defined for the present purpose as the average of per adult national income in Germany, France and the United Kingdom). Using purchasing power parity estimates (PPP), we find that average income in the Middle East stood at about 70-75% of the European average in 1990. It then fell during the 1990s

and early 2000s, down to about 60% around 2003-2004, and finally rose back to about 65-70% between 2004 and 2016. Using market exchange rates (MER), the ratio has also been stagnating over the 1990-2016 period, but at substantially lower levels, i.e. around 25-30% of the West European average rather than 60-70% (see Figure 2.2).

In our view, both the PPP and the MER viewpoints express valuable and complementary aspects of international inequality patterns. The PPP viewpoint should of course be preferred if we are interested in the living standards of the inhabitants living, working and spending their incomes in the various countries (which is the case of most people). However the MER viewpoint is more relevant and meaningful if we are interested in external economic relations: e.g. the ability of tourists and visitors from Europe or from Gulf countries when they travel to other countries; or the ability of migrants or prospective migrants from Egypt or Syria to send part of their euro wages back home. Here market exchange rates matter, and may also play an important role on the perceptions of inequality. Whatever the viewpoint, it is important to have in mind that per adult average income benefited from very little growth over the 1990-2016: in effect, the vast majority of aggregate national income growth was absorbed by the rise of population (see Figure 2.3).

Next, and most importantly, it is critical to stress that there exists enormous and persistent between-country inequality behind the Middle East average. In order to summarize the changing population and income structure of the Middle East, it is helpful to decompose the region into five blocs: (i) Turkey; (ii) Iran; (iii) Egypt; (iv) Iraq and Syria and other non-Gulf countries: Jordan, Lebanon, Palestine, Yemen); and (v) Gulf countries (including Saudi Arabia, Oman, Bahrain, UAE, Qatar and Kuwait) (see Table 1 above). Each of the first four blocs represents about 20-25% of total population of the Middle East, with relatively little variations over the 1990-2016 period (except for a slight rise in the share of the Iraq-Syria-other bloc). The main change in the structure of Middle East population over the past quarter of a century is the rise of the population share

of Gulf countries, from about 10% in 1990 to 15% in 1996 (Figure 2.4). This is almost entirely due to the rise of migrant workers in oil-rich countries (see below).

If we now look at average income patterns in these five sub-regions, we find that per adult national income is substantially below average everywhere except in Gulf countries (see Figures 2.5-2.6). One can distinguish between two groups: Turkey and Iran, where average incomes have generally been around 50-60% of the West European average in PPP terms (with a significant rise of Turkish incomes over the 2001-2015 period, in contrast to Iranian stagnation); and Egypt and Iraq-Syria-other, where average incomes have always stood at significantly lower levels (around 30-40% of West European average in PPP terms) (see Figure 2.5). Using market exchange rates, we find that Egypt-Iraq-Syria-other have stagnated around 10-15% of the West European average (see Figure 2.6). It is also worth noting that Turkey rises slightly above Middle East average when we consider MER series, while Iran falls toward the levels of Egypt and Iraq-Syria, reflecting the weakness of the Iranian currency and the relative strength of the Turkish lira.

As compared to the rest of Middle East, Gulf countries clearly belong to a different category. In PPP terms, their average per adult national income was about three times that of Western Europe in 1990, and almost two times in 2016; in MER terms, their income was 40% higher than the West European level in 1990, and is currently about 10% lower (see Figure 2.7). In brief: the enormous gap in average incomes between Gulf countries and the more populated Middle East countries has been trending downward in the past 25 years. However, two remarks are in order. First, the income gap is still enormous: Gulf countries represent only 15% of the Middle East population in 2016, but they receive between 42% (in PPP terms) and 47% (in MER terms) of total Middle East income (Table 2.1). Back in 1990, their population share was 10%, and their income share was between 44% (PPP) and 48% (MER). Next, the fall in the income gap between Gulf countries and the rest of the Middle East reflects a number of complex and contradictory forces. It is partly due to the evolution of oil prices and output levels, as well as to

the relative fast output growth in non-Gulf countries like Turkey. But it is also due to the very large rise of migrant workers, and the consequently migration-led reduction of per adult national income in Gulf countries: the massive inflow of foreign workers (especially in the construction sector and domestic services sector) resulted in a stronger increase in the population denominator than in the income numerator of Gulf countries. By putting together census and survey data for the various countries, we find that the overall rise of the population share of Gulf countries (from 10% to 15% of total Middle East population) is almost entirely due to the massive rise in foreign workers, which increased from less than 50% in 1990 to almost 60% of the total population in 2016 (Figure 2.8). From this viewpoint, it is also helpful to distinguish between two groups of Gulf countries: one group made of Saudi Arabia, Oman and Bahrain, where nationals still make a (small) majority of the population (the foreign population share has been relatively stable around 40-45% of total adult population between 1990 and 2016); and another group made of United Arab Emirates (UAE), Kuwait and Qatar, where the nationals make a smaller and smaller minority of the resident population (the foreign share rose from 80% to 90%; see Figure 2.8). This second group made about one quarter of total population of Gulf countries in 1990, up to about one third by 2016.

Finally, it is worth stressing that our ability to measure income inequalities within oil-rich countries is relatively limited. By exploiting available household surveys, we find that the ratio between average per adult survey income of nationals and foreigners is particularly large in UAE-Kuwait-Qatar (which is not surprising, given the very small share of nationals), and most importantly that this ratio has increased over time, from 250% in 1990 to around 350% in 2016 (see Figure 2.9). In Gulf countries where the national-vs-foreigner population structure is closer to 50-50 (i.e. Oman-Bahrain), the income ratio between nationals and foreigners appears to be less extreme (but still substantial: around 160%). These estimates are solely based upon self-reported survey data (with no correction for the under-estimation of top incomes), and should therefore be

considered as a lower bound. In addition, we are not able to include Saudi Arabia (by far the most populated among Gulf countries) in these computations due to a lack of access to adequate survey data.

One should also emphasize that the fraction of national income covered by household surveys is particularly low in Gulf countries (about 30%, vs 40-50% in other Middle East countries; see Table 2.2), i.e. a very significant fraction of the national income is not properly attributed to households. To the extent that nationals benefit from the excluded income components (which typically refer to the undistributed profits of oil corporations and the accumulated capital income of sovereign wealth funds) more than foreigners, this may also contribute to under-estimating the nationals-vs-foreigners income ratio.

2.4.2 *Extreme Level of Income Concentration in the Middle East*

We now present our main results regarding the level of income concentration in the Middle East. According to our benchmark estimates, the share of total income going to top 10% income earners is about 61% in the Middle East, as compared to 36% in Western Europe and 47% in the USA (Figure 2.10).

Several remarks are in order. First, these three regions have comparable population sizes (with a total population of about 410 million in the Middle East, 420 million in Western Europe, and 320 million in the USA) and a relatively large degree of cultural, historical and linguistic proximity, so we feel that the comparison is legitimate and meaningful.

Next, the fact that we find much higher inequality levels in the Middle East appears to be extremely robust. We obtain the same finding not only in the benchmark series, but also in all variant series, often with a larger margin. Also, we focus on Figures 2.10 - 2.13 on the latest years available (2012-2016), and as we shall see below, the inequality gap with other regions was if anything even higher in previous decades (see section 4.3). Most importantly, we stress again that our inequality estimates for the Middle East are

based upon highly conservative estimates of within-country inequality. In particular, it is very likely that we severely under-estimate the true level of inequality within Gulf countries. With better data sources, it is plausible that we would find even higher inequality levels (e.g. with a top decile income share around 65%-70% in the Middle East, or even higher).

Income inequality also appears to be significantly higher in the Middle East than in Brazil - a country with population around 210 million that is often described as one of the most unequal in the world, and where the top decile income share is about 55% (according to the recent DINA estimates constructed by [Morgan \(2017\)](#)). The only country for which we find slightly higher inequality estimates than for the Middle East is South Africa, with about 62% for the top decile income share for the latest available years (Alvaredo and Atkinson, 2010, and series updated in 2017 in WID.world).

It is worth stressing that the origins of inequality are obviously very different in these different groups of countries. In the case of the Middle East, they are largely due to the geography of oil ownership and the transformation of oil revenues into permanent financial endowments. In contrast, extreme inequality in South Africa is intimately related to the legacy of the Apartheid system: until the early 1990s, only the white minority (about 10% of the population, which until today roughly corresponds to the top 10% income group) had full mobility and ownership rights. In Brazil, the legacy of racial inequality also plays an important role (it was the last major country to abolish slavery in 1887, at a time when slaves made up about 30% of the population), together with huge regional inequalities. It is striking to see that the Middle East, in spite of its much larger racial and ethno-cultural homogeneity, has reached inequality levels that are comparable to - or even higher than - those observed in South Africa or Brazil.

It is also worth stressing that inequality levels in the Middle East appear to be significantly larger than those observed in giant countries with much bigger populations such as China and India (see Figures [2.10- 2.11](#)). Here we use inequality estimates that were

recently constructed for China by [Piketty et al. \(2017\)](#). These estimates are obviously far from perfect, but they probably tend to minimize the gap with the Middle East (in particular, we have access to more extensive income tax data for China and India than the for the Middle East).

In brief: according to our estimates, the Middle East appears to be the most unequal region in the world (possibly with the exception of South Africa). This is true for the top decile income share, as well as for other inequality indicators; e.g. the top percentile income share is about 27% in the Middle East, vs. 12% in Western Europe, 20% in the USA, 28% in Brazil, 18% in South Africa, 14% in China and 21% in India (see [Figure 2.11](#)).¹² This is also true for synthetic indicators such as the Gini coefficient (see on-line series).

We find it particularly informative to compare the overall levels of the income shares going to the bottom 50%, the middle 40% and the top 10% and 1% in the Middle East and other countries (see [Figures 2.12-2.13](#)). For instance, according to our benchmark estimates, the bottom 50% of the population receives about 9% of total income in the Middle East (vs. 18% in Europe), as compared to 61% for the top 10% (vs. 36% in Western Europe). This clearly illustrates that differences in distributions can make an enormous difference when comparing income and welfare levels across countries.

2.4.3 Analyzing the evolution of income inequality in the Middle East and other robustness checks

We now turn to our results regarding the evolution of income inequality in the Middle East over the 1990-2016 period. We should stress again that the data sources at our disposal are insufficient to properly analyze trends in inequality. In our benchmark

12. Note that South Africa is more unequal than Brazil or the Middle East in terms of top 10% income share, but less unequal in terms of top 1% income share. This seems to reflect the existence of a 10% elite (to a large extent the white population) that is much richer than the rest of the population and relatively homogenous (as compared to the graduation between the top 10% and the top 1% prevailing in other countries). Note also that our limited data sources on very top earners in the Middle East make it very difficult to provide a proper comparison with Brazil regarding the top 1% share.

estimates, we find a declining inequality trend at the regional level between 1990 and 2010, followed by a rising trend between 2010 and 2016. However these are trends of relatively small magnitude, and it is unclear whether these are robust findings (more on this below). As a first order approximation, our main finding - and probably the only robust finding - is that income concentration is very high and approximately constant in the Middle East region (taken as a whole) over the 1990-2016 period.

According to our benchmark estimates, the top 10% income share fluctuates around 60%-65% of total income between 1990 and 2016, while the bottom 50% income share fluctuates around 8%-10% of total income (see Figure 2.14). We have constructed a large number of variant estimates, and these orders of magnitude appear to be robust. When we move from market-exchange-rate estimates (which we use as benchmark series) to purchasing-power-parity estimates, inequality levels decline a little bit - as one might expect (see variant 1 in Figure 2.15).¹³ When we change the geographical definition of the Middle East, for instance by excluding Turkey (a country whose average income is intermediate between the poorest countries - Egypt, Iraq, Syria, Yemen, etc. - and the oil-rich Gulf countries, and which therefore to moderate inequality at the regional level), we find somewhat higher inequality levels (see variant 2 in Figure 2.15). But in any case the effect is not very large, especially as compared to the difference in inequality levels between the Middle East and other world regions.

As one can see from Figures 2.16-2.17 where we compare the evolution of the top 10% and top 1% income shares in the Middle East and other world regions between 1990 and 2016, the striking fact is that income inequality has always been much higher in the Middle East.¹⁴ The inequality gap was particularly large in 1990 (in a way, the Middle East has been a pioneer region in terms of extreme inequality). The gap declined during

13. We tend to prefer MER estimates because they are in a way more comparable to those estimated for other world regions (i.e. we do not use price differentials when estimating income inequality within the USA, Brazil, China or India). But as noted above both perspectives offer valuable and complementary insights.

14. The same conclusion holds true when we look at other inequality indicators such as the bottom 50% income share or the Gini coefficient.

the 1990s-2000s, as other world regions gradually became more and more unequal (and as Middle East inequality declined a little bit). But the gap remains substantial, and the Middle East continues to be in 2016 the most unequal region in the world (with the possible exception of South Africa).¹⁵

In order to better understand the origins of our high inequality estimates, we compare on Figures 2.18-2.20 the results obtained with our benchmark national income series (combining survey data, national accounts, income tax and wealth data), the results obtained with the fiscal income series (ignoring the wealth correction, which as one can see plays a relatively minor role) and the results obtained with the survey data alone. Here we distinguish between the survey income series obtained with different country-level average incomes (anchored on observed country-level per adult national incomes) and the survey income series simulated by assuming fixed country-level average incomes (thereby neutralizing the impact of between-country inequality). As one see, both the within-country inequality effect (fiscal data correction) and the between-country inequality effect play an important role in accounting for the very high inequality estimates.

Finally, we have also simulated what would have been the evolution of income inequality in the Middle East over the 1990-2016 if within-country inequality had remained fixed at the observed 1990 level. As one can see from Figure 2.21, the evolution of total inequality at the level of the Middle East taken as a whole would have been virtually the same, with declining inequality between 1990 and 2010 and rising inequality since 2010. This shows that our estimates are mostly driven by the evolution of between-country inequality. This is partly due to the fact that we do not have survey data for all years (see Table 2.2 above), so that for some countries our inequality estimates display very little time variations (and in some cases no time variation at all). This is also due to the fact that even in countries with multiple survey years, we observe limited variations in income inequality, and these variations tend to compensate each other.

15. Unfortunately available series for the top 10% and top 1% share in South Africa do not cover all years, so it is difficult to make a complete comparison with the Middle East at this stage.

E.g. household surveys indicate that income inequality declined somewhat in Turkey between 2003 and 2007, but then rose between 2007 and 2016; inequality increased in Lebanon between 2005 and 2008 and then stabilized; inequality declined in Egypt between 1999 and 2010, but then rose between 2010 and 2015; inequality increased in the UAE between 1998 and 2009 according to the top 10% income share (but declined according to the top 1% income share); inequality increased according to both indicators in Qatar between 2007 and 2012; and so on (see Table 2.3).

Most of these variations are relatively modest in magnitude, so it is not surprising that most of the evolution of Middle East inequality is driven by the evolution of between-country inequality and the fact that the gap in average income between oil-rich countries and other countries has been trending downwards (but is still very large in level). If we had access to adequate income tax data throughout the 1990-2016 period, we might reach different conclusions and find a strong within-country rising inequality trend (such as the one found in a large number of very different countries across the world, e.g. in the USA, Europe, India, China, South Africa, Russia, with varying magnitudes). It is also possible that Middle East countries - like Brazil - belong to a different category, i.e. countries where inequality has always been very large historically (so that it did not rise in recent decades). Given the data sources at our disposal, we are not able to conclude with a satisfactory degree of precision.

2.5 *Conclusion*

In this paper we have combined household surveys, national accounts, income tax data and wealth data in order to estimate the level and evolution of income concentration in the Middle East for the period 1990-2016. According to our benchmark series, the Middle East appears to be the most unequal region in the world, with a top decile income share as high as 61%, as compared to 36% in Western Europe, 47% in the USA, and 55% in Brazil. This is due both to enormous inequality between countries (particularly between oil-rich and population-rich countries) and to large inequality within countries (which we probably under-estimate, given the limited access to proper fiscal data). To our knowledge, this is the first attempt to measure income inequality at the level of the Middle East taken as a whole, and also the first attempt to use income tax data and generalized Pareto interpolation techniques to correct household survey data in the Middle East.

The data at our disposal is highly imperfect, and we still face considerable limitations in our ability to measure inequality in the Middle East. In particular, there is much uncertainty about inequality trends in the period under study. However the general conclusion that the overall inequality level is one of the highest in the world - if not the highest in the world, with the possible exception of South Africa - appears to be very robust.

In conclusion, we would like to stress the importance of increasing transparency on income and wealth in the Middle East. In particular, it is critical that Middle East countries provide access to household surveys micro-files, and even more importantly that they provide access to income tax data (at least in the form of income tax tabulations). It is very difficult to have an informed public debate about inequality trends - and also about a large number of substantial policy issues such as taxation and public spending - without proper access to such data. While the lack of transparency on income and

wealth is an important issue in many - if not most - areas of the world, it appears to be particularly extreme in the Middle East, and arguably raises in itself a problem of democratic accountability, quite independently from the actual level of inequality.

Finally, our results regarding the enormous level of income inequality in the Middle East region naturally point toward the need to develop mechanisms of regional redistribution and investment. In a way, this is already happening, in the sense that oil-rich countries regularly make loans to poorer countries (e.g. Saudi Arabia to Egypt), and that these loans sometimes include implicit or explicit subsidies. However such mechanisms are usually of limited magnitude, and tend to be highly unpredictable. Given the enormous concentration of gross domestic product and national income in the region, mechanisms of regional investment funds similar to those developed in the European Union (with permanent transfers between the richest and the poorest countries of the order of several percentage points of GDP) could make a large difference. These issues would deserve more attention in future research.

Figure 2.1 – Per adult national income: Middle East vs W. Europe 1990-2016

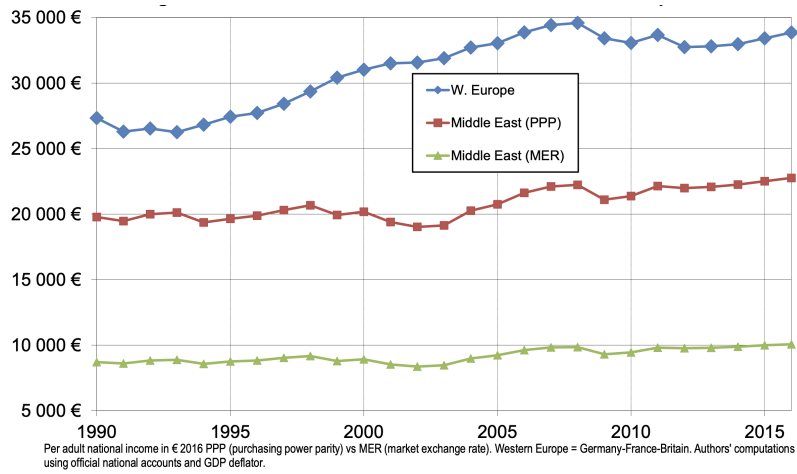


Figure 2.2 – Per adult national income: ratio Middle East /W.Europe

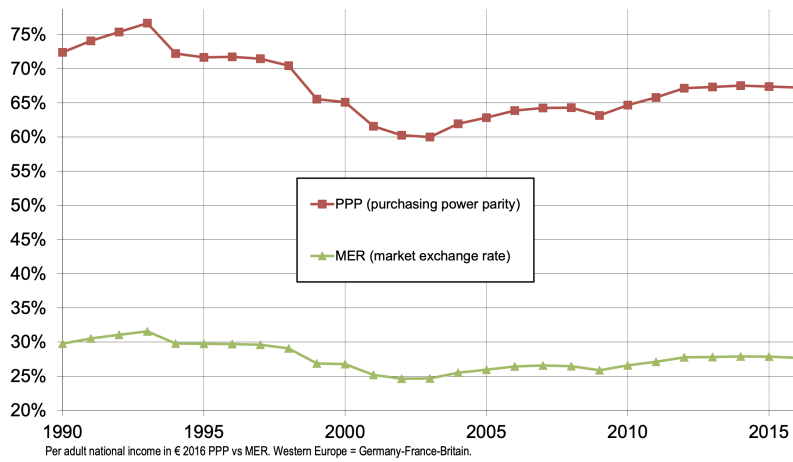


Figure 2.3 – Cumulated growth in the Middle East: income vs population

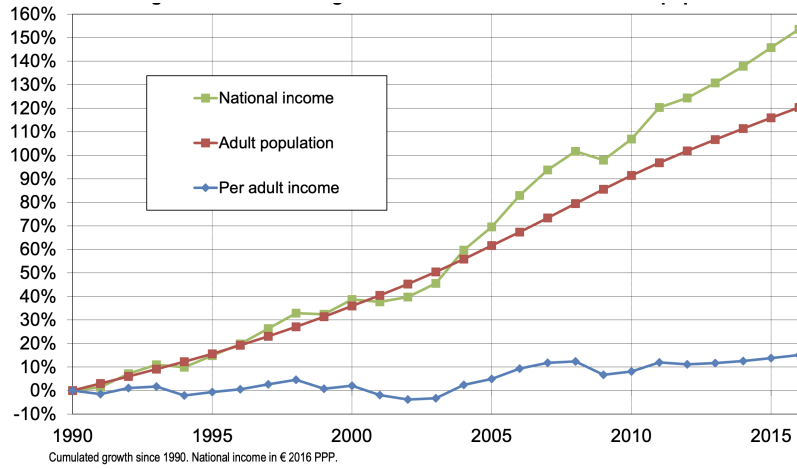


Figure 2.4 – Population shares in the Middle East, 1990-2016

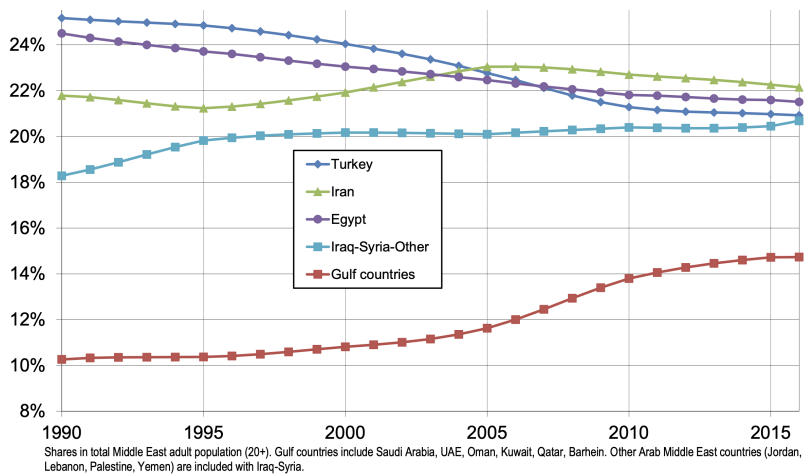


Figure 2.5 – Per adult national income: ratio Middle East /W.Europe (PPP)

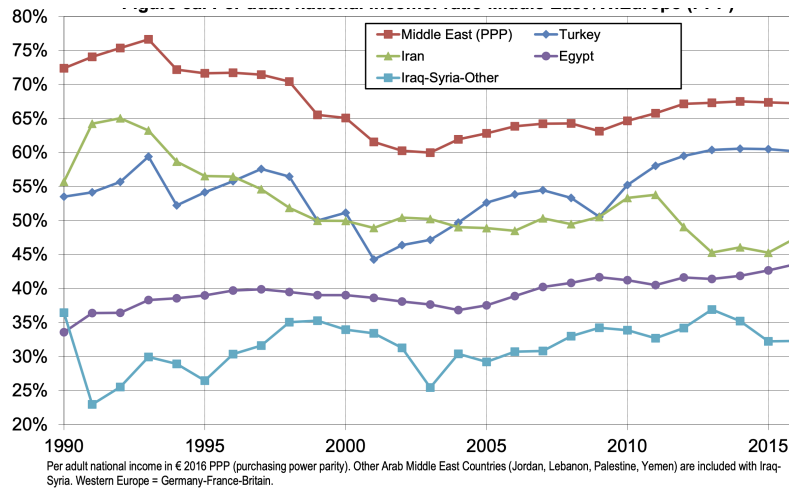


Figure 2.6 – Per adult national income: ratio Middle East /W.Europe (MER)

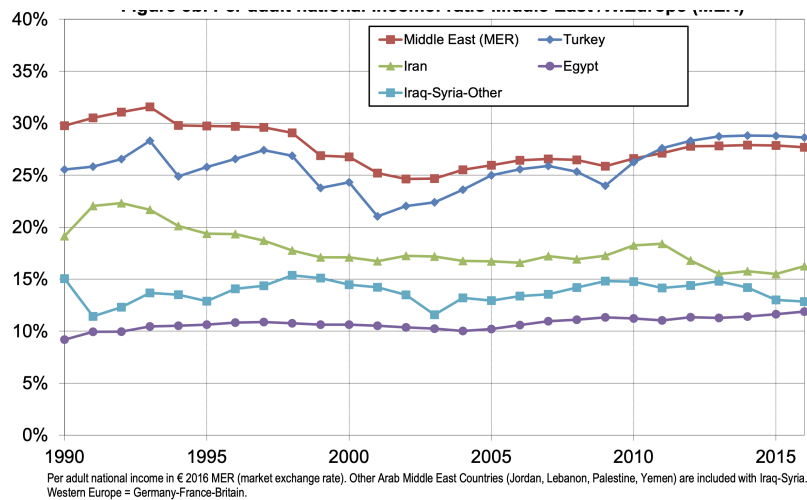


Figure 2.7 – Per adult national income: ratio Gulf countries/W. Europe

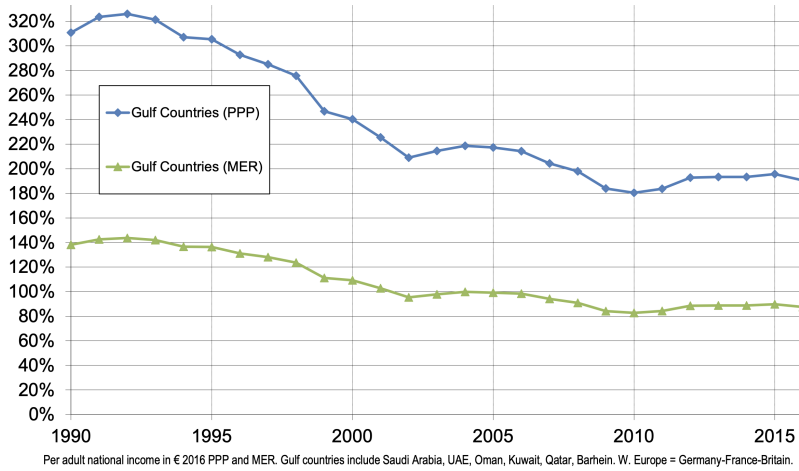


Figure 2.8 – Shares of foreigners in Gulf Countries, 1990-2016

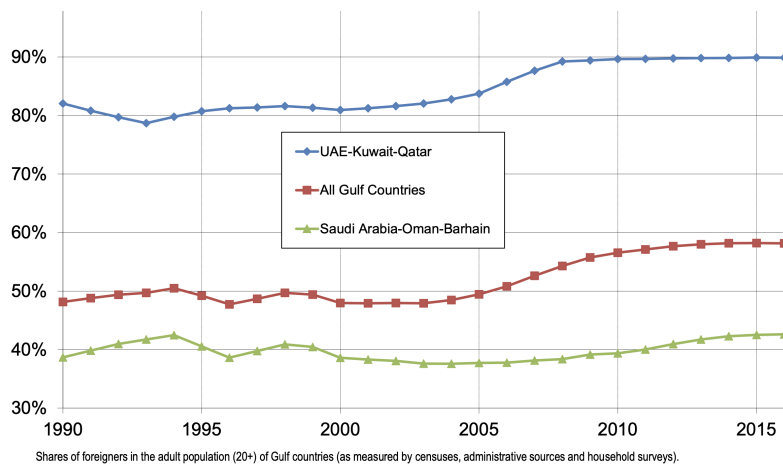


Figure 2.9 – Income ratios Nationals/Foreigners in Gulf countries, 1990-2016

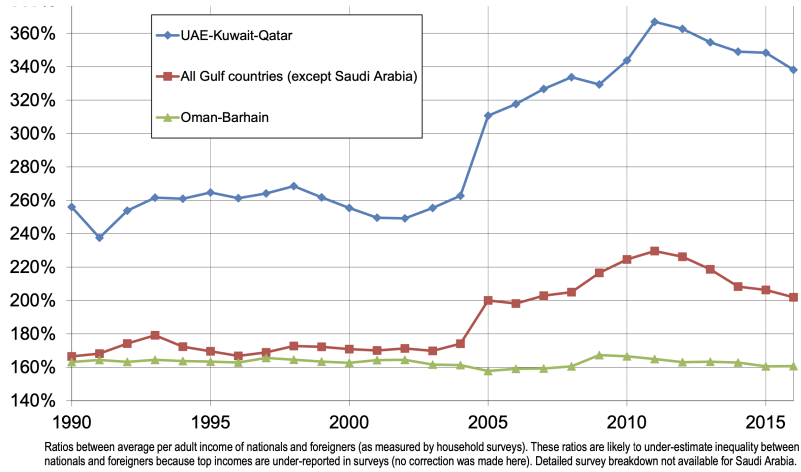


Figure 2.10 – Top 10% income share: Middle East vs other countries

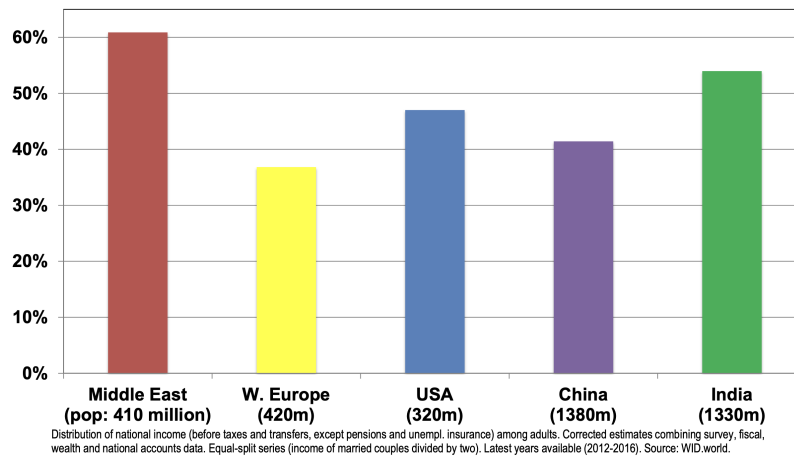
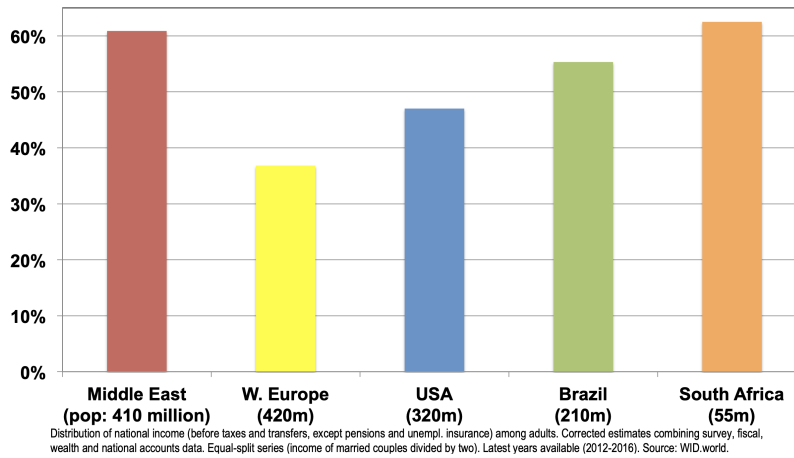


Figure 2.11 – Top 1% income share: Middle East vs other countries

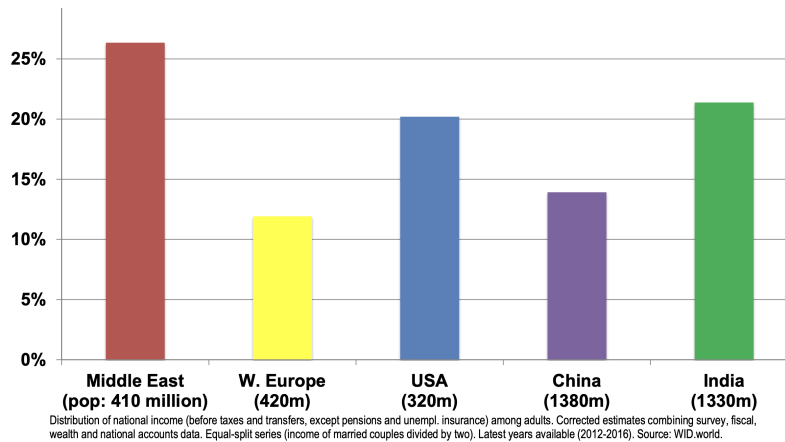
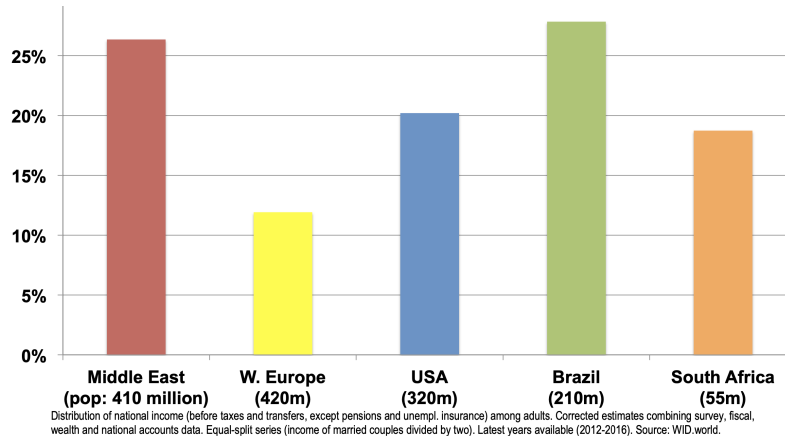


Figure 2.12 – Bottom 50% vs Middle 40% vs Top 10% income shares

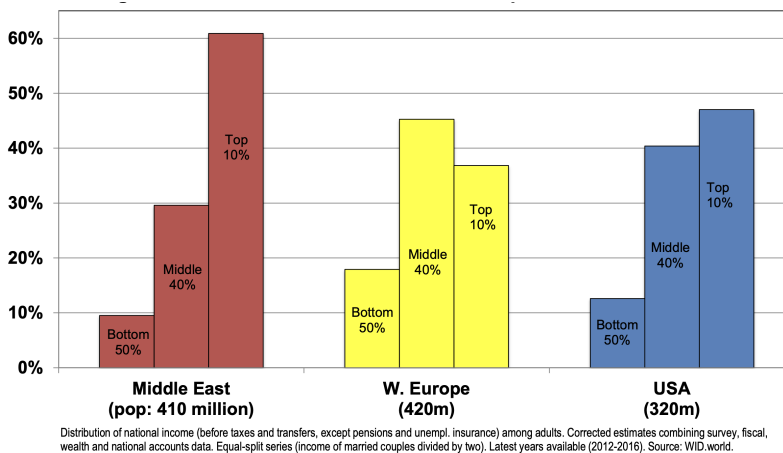


Figure 2.13 – Bottom 50% vs Top 1% income shares

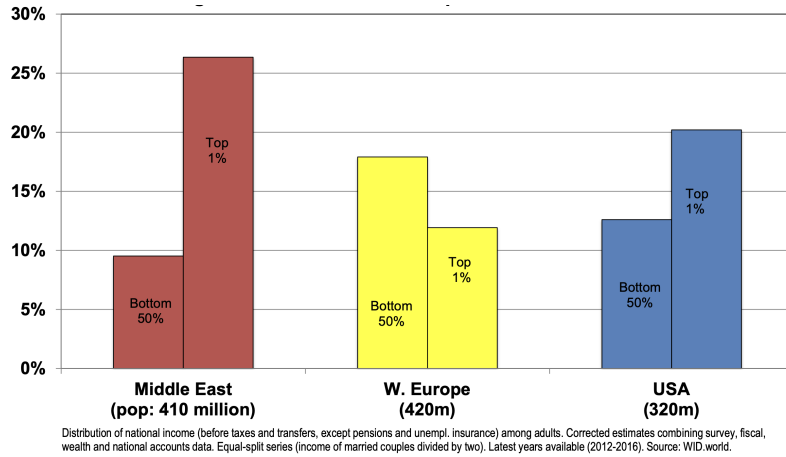


Figure 2.14 – Income shares in the Middle East, 1990-2016 (benchmark series)

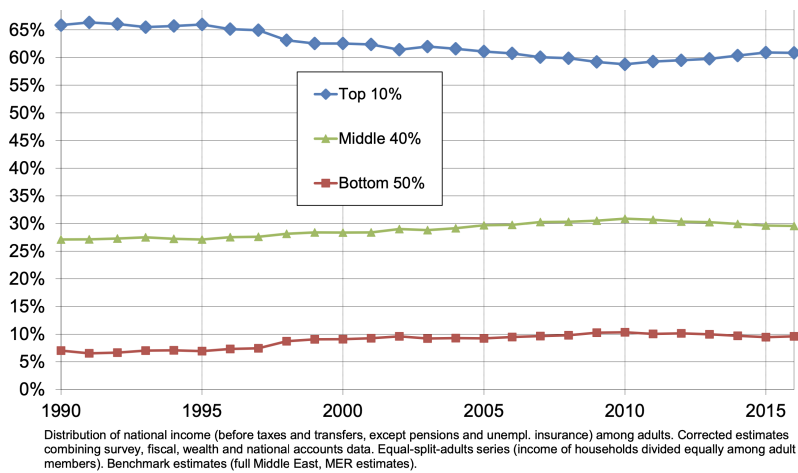


Figure 2.15 – Top income shares in the Middle East, 1990-2016 (variants)

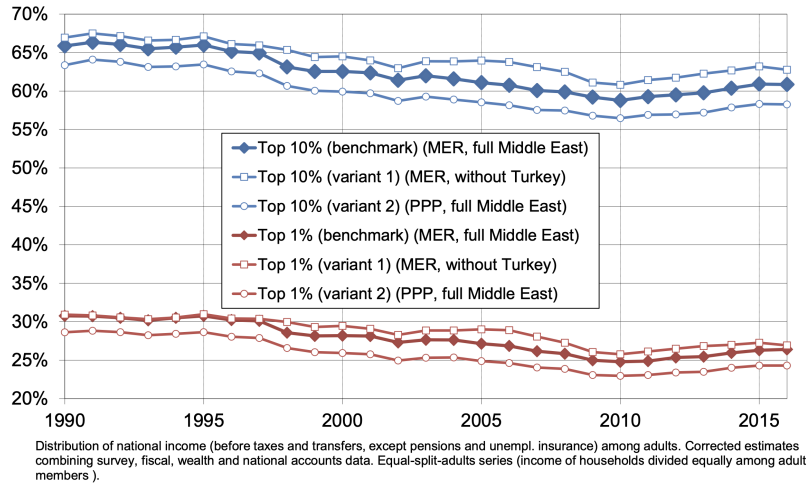


Figure 2.16 – Top 10% income share 1990-2016: Middle East vs other countries

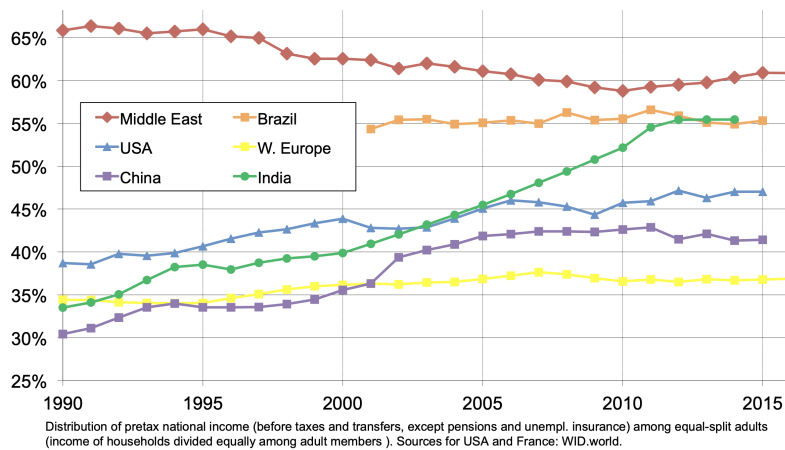


Figure 2.17 – Top 1% income share 1990-2016: Middle East vs other countries

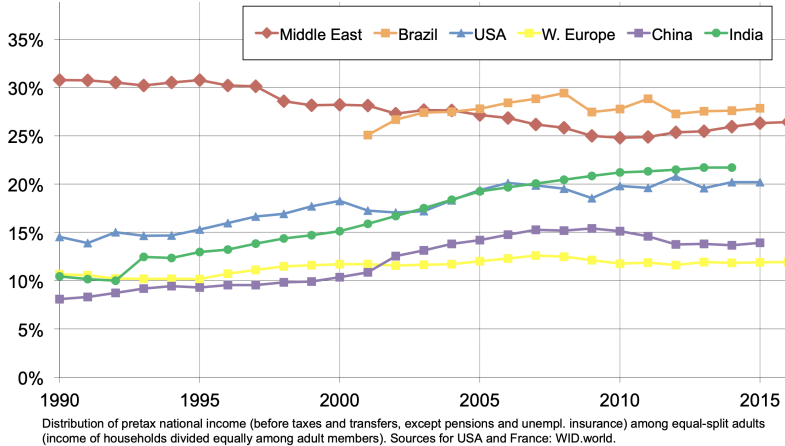


Figure 2.18 – Decomposing the level of Middle East top 10% income share

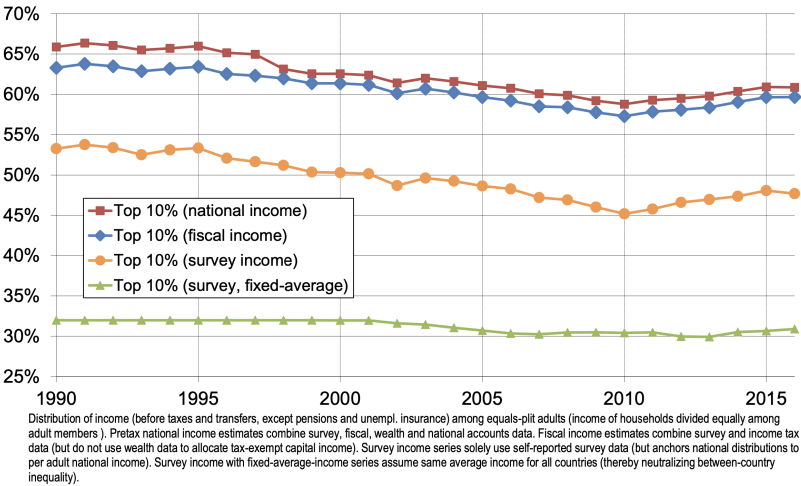


Figure 2.19 – Top 1% income share in the Middle East, 1990-2016

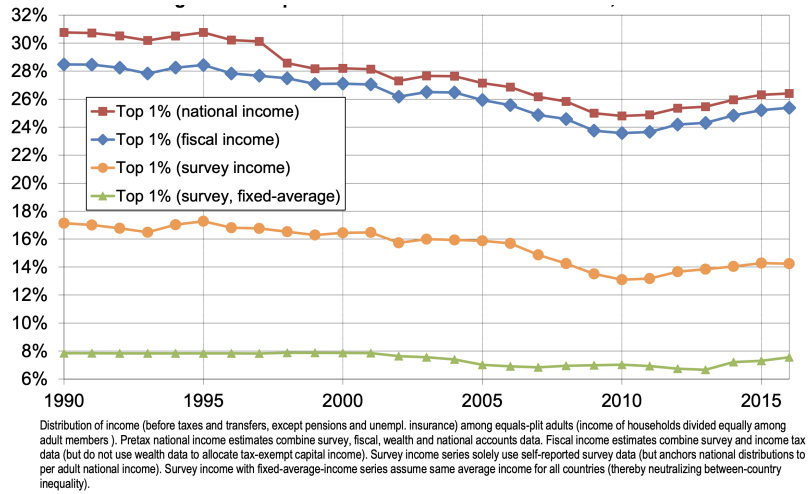


Figure 2.20 – Decomposing Gini coefficients in the Middle East, 1990-2016

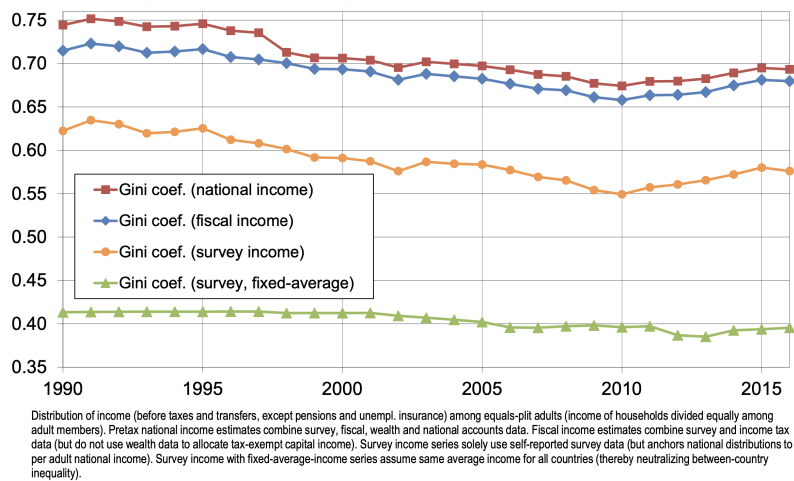


Figure 2.21 – Decomposing the evolution of Mid. East top 10% income share

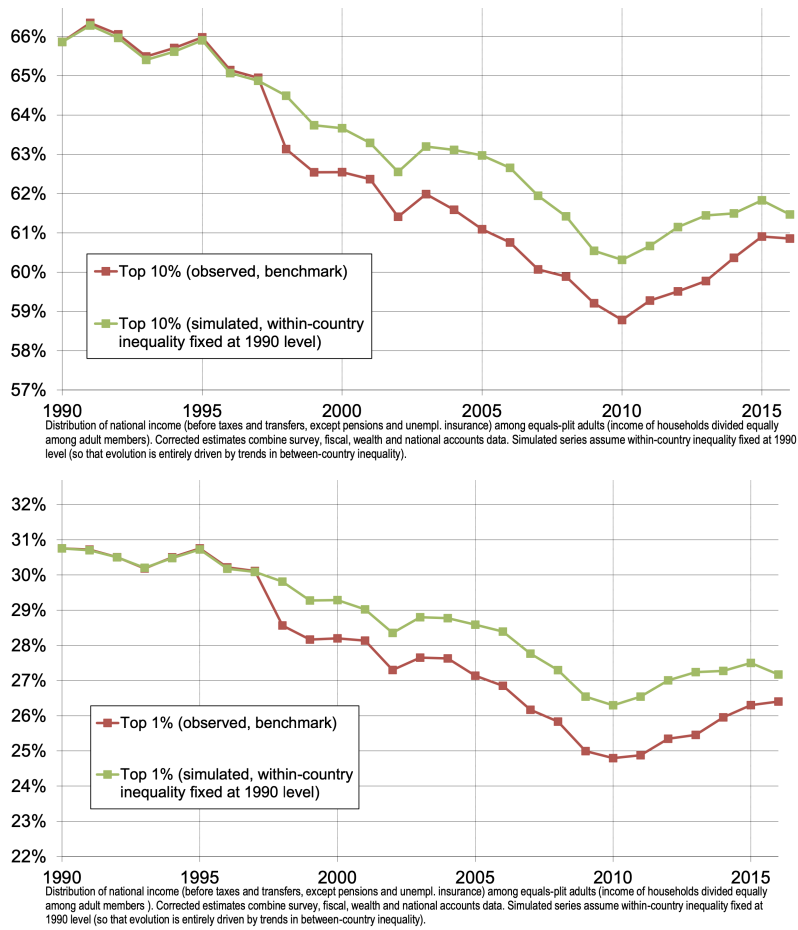


Table 2.1 – Population and income in the Middle East (2016)

	Population (million)	Adult Population (million)	Adult population (% of ME total)	National Income (Billion PPP Euro 2016)	% ME Total Income (PPP)	National Income (Billion MER Euro 2016)	% ME Total Income
Turkey	80	53	21 %	1073	19 %	548	22 %
Iran	80	56	22 %	896	16 %	330	13 %
Egypt	93	54	22 %	800	14 %	234	9 %
Iraq-Syria-Other (non-Gulf)	102	52	21 %	570	10 %	243	10 %
Iraq	38	18	7 %	354	6 %	112	4 %
Syria	19	10	4 %	47	1 %	28	1 %
Jordan	8	4	2 %	57	1 %	30	1 %
Lebanon	6	4	2 %	57	1 %	40	2 %
Palestine	5	2	1 %	16	0 %	12	0 %
Yemen	27	13	5 %	39	1 %	21	1 %
Gulf Countries	54	37	15 %	2394	42 %	1179	47 %
Saudi Arabia	32	20	8 %	1313	23 %	575	23 %
Oman	5	3	1 %	118	2 %	47	2 %
Bahrain	1	1	0 %	46	1 %	26	1 %
UAE	9	8	3 %	430	7 %	283	11 %
Kuwait	4	3	1 %	258	5 %	122	5 %
Qatar	2	2	1 %	229	4 %	126	5 %
Total Middle East	409	252	100 %	5733	100 %	2534	100 %

Table 2.2 – Household surveys used in this paper (1990-2016)

	Survey years	Average ratio (total survey income)/(national income)
Turkey	1994, 2002-2016	43 %
Iran	2010, 2013	49 %
Egypt	1999, 2004, 2008, 2010, 2012, 2015	40 %
Iraq-Syria-Other (non-Gulf)	1992-2013	53 %
Iraq	2007	60 %
Syria	2004	56 %
Jordan	1992, 2002, 2006, 2008, 2010, 2013	70 %
Lebanon	2007	37 %
Palestine	1996-1998, 2004-2008, 2010-2011	65 %
Yemen	2006	33 %
Gulf Countries	1995-2013	30 %
Saudi Arabia	2008	30 %
Oman	2010	29 %
Bahrain	1995, 2005, 2015	37 %
UAE	1998, 2009	39 %
Kuwait	2007, 2013	21 %
Qatar	2007, 2012	23 %

Table 2.3 – Population and income in the Middle East (2016)

Country	Year	Bottom 50%	Middle 40%	Top 10%	Top 1%	Gini	P10/ average	P50/ average	P90/ average	P99/ average	Pareto b(10%)	Pareto b(50%)	Pareto b(90%)
Bahrain	1995	14 %	32 %	54 %	19 %	61 %	18 %	46 %	175 %	836 %	6,2	3,8	2,9
	2005	15 %	33 %	52 %	18 %	59 %	20 %	49 %	177 %	842 %	5,5	3,5	2,8
	2015	14 %	33 %	53 %	18 %	60 %	19 %	47 %	178 %	856 %	5,6	3,6	2,8
Egypt	1999	17 %	32 %	51 %	19 %	56 %	25 %	50 %	165 %	814 %	4,3	3,3	2,9
	2004	18 %	33 %	49 %	18 %	55 %	26 %	53 %	169 %	783 %	4,2	3,1	2,7
	2008	18 %	34 %	49 %	19 %	54 %	27 %	53 %	167 %	761 %	4,1	3,1	2,7
	2010	19 %	35 %	46 %	17 %	52 %	28 %	56 %	170 %	713 %	3,9	2,9	2,6
	2012	19 %	35 %	46 %	16 %	52 %	28 %	57 %	170 %	704 %	3,9	2,9	2,6
	2015	18 %	33 %	49 %	19 %	54 %	27 %	54 %	163 %	762 %	4,0	3,0	2,8
Iran	2010	14 %	35 %	51 %	18 %	59 %	17 %	51 %	183 %	779 %	6,4	3,4	2,6
	2013	17 %	35 %	48 %	16 %	55 %	22 %	54 %	180 %	748 %	4,9	3,1	2,5
Iraq	2007	15 %	32 %	53 %	22 %	59 %	20 %	50 %	165 %	823 %	5,4	3,4	3,0
Jordan	1992	15 %	33 %	51 %	20 %	58 %	21 %	50 %	170 %	780 %	5,3	3,4	2,8
	2002	18 %	36 %	46 %	15 %	53 %	25 %	56 %	182 %	674 %	4,3	2,9	2,4
	2006	17 %	36 %	47 %	15 %	54 %	23 %	55 %	186 %	700 %	4,8	3,0	2,4
	2008	18 %	35 %	47 %	16 %	54 %	25 %	56 %	178 %	707 %	4,3	3,0	2,5
	2010	17 %	32 %	52 %	22 %	57 %	22 %	52 %	158 %	816 %	4,9	3,2	3,1
	2013	17 %	35 %	48 %	16 %	54 %	23 %	55 %	181 %	710 %	4,8	3,0	2,5
Kuwait	2007	13 %	33 %	54 %	18 %	62 %	19 %	41 %	196 %	846 %	5,9	4,2	2,6
	2013	13 %	33 %	54 %	18 %	62 %	18 %	38 %	213 %	800 %	6,0	4,6	2,4
Lebanon	2005	13 %	35 %	52 %	22 %	62 %	15 %	48 %	180 %	760 %	7,2	3,6	2,7
	2006	13 %	35 %	52 %	21 %	61 %	16 %	49 %	180 %	751 %	7,1	3,5	2,7
	2007	11 %	33 %	57 %	24 %	66 %	12 %	42 %	184 %	847 %	9,5	4,2	2,9
	2008	11 %	32 %	57 %	24 %	66 %	11 %	42 %	184 %	856 %	9,7	4,3	2,9
	2009	11 %	33 %	57 %	23 %	66 %	12 %	42 %	185 %	855 %	9,5	4,2	2,9
	2010	11 %	32 %	57 %	23 %	66 %	12 %	42 %	185 %	863 %	9,6	4,3	2,9
	2011	11 %	32 %	57 %	23 %	66 %	11 %	41 %	185 %	867 %	9,7	4,3	2,9
	2012	10 %	32 %	58 %	24 %	67 %	11 %	41 %	186 %	874 %	9,8	4,4	2,9
	2013	11 %	32 %	57 %	23 %	67 %	11 %	41 %	186 %	873 %	9,7	4,4	2,9
	2014	11 %	32 %	57 %	23 %	66 %	12 %	42 %	185 %	861 %	9,6	4,3	2,9
Oman	2010	10 %	35 %	55 %	18 %	66 %	8 %	43 %	201 %	853 %	13,4	4,2	2,6
Palestine	1996	15 %	37 %	48 %	14 %	57 %	17 %	54 %	201 %	701 %	6,3	3,2	2,3
	1997	15 %	37 %	48 %	14 %	57 %	17 %	53 %	201 %	706 %	6,4	3,2	2,3
	1998	15 %	37 %	49 %	14 %	58 %	17 %	53 %	200 %	714 %	6,6	3,2	2,3
	2004	15 %	36 %	49 %	15 %	58 %	16 %	53 %	191 %	751 %	6,8	3,2	2,4
	2005	14 %	37 %	50 %	16 %	59 %	15 %	51 %	203 %	743 %	7,4	3,4	2,3
	2006	14 %	36 %	50 %	15 %	59 %	16 %	51 %	193 %	761 %	7,0	3,4	2,5
	2007	13 %	35 %	52 %	17 %	62 %	14 %	47 %	198 %	790 %	8,1	3,7	2,5
	2008	14 %	36 %	50 %	14 %	59 %	16 %	50 %	199 %	739 %	7,0	3,5	2,4
	2010	13 %	35 %	52 %	17 %	61 %	14 %	48 %	196 %	782 %	7,6	3,7	2,5
	2011	13 %	36 %	51 %	16 %	60 %	15 %	49 %	198 %	745 %	7,3	3,5	2,4
	Qatar	2007	12 %	36 %	52 %	17 %	62 %	14 %	47 %	196 %	824 %	8,1	3,7
2012		14 %	33 %	53 %	19 %	61 %	18 %	47 %	180 %	881 %	6,2	3,7	2,8
SaudiArabia	2008	13 %	34 %	53 %	16 %	61 %	18 %	44 %	201 %	795 %	6,2	3,9	2,5
Syria	2004	14 %	37 %	49 %	15 %	58 %	16 %	51 %	201 %	624 %	6,7	3,3	2,3
Turkey	1994	8 %	31 %	61 %	28 %	71 %	7 %	36 %	165 %	970 %	15,4	5,1	3,5
	2002	14 %	31 %	55 %	22 %	62 %	17 %	45 %	169 %	909 %	6,3	3,8	3,1
	2003	14 %	31 %	55 %	22 %	61 %	18 %	46 %	167 %	910 %	6,1	3,8	3,1
	2004	14 %	32 %	53 %	21 %	60 %	18 %	48 %	173 %	815 %	6,1	3,6	2,9
	2005	15 %	34 %	51 %	19 %	58 %	19 %	51 %	177 %	797 %	5,9	3,3	2,7
	2006	16 %	34 %	50 %	18 %	57 %	21 %	52 %	176 %	793 %	5,3	3,2	2,7
	2007	16 %	35 %	49 %	17 %	56 %	21 %	53 %	176 %	751 %	5,2	3,1	2,6
	2008	16 %	34 %	50 %	18 %	57 %	20 %	53 %	173 %	748 %	5,4	3,2	2,7
	2009	15 %	33 %	52 %	19 %	58 %	20 %	51 %	171 %	858 %	5,6	3,3	2,8
	2010	16 %	33 %	51 %	20 %	58 %	21 %	51 %	171 %	793 %	5,3	3,3	2,8
	2011	16 %	33 %	51 %	19 %	58 %	20 %	50 %	173 %	811 %	5,4	3,4	2,8
	2012	16 %	33 %	51 %	20 %	58 %	21 %	50 %	172 %	772 %	5,2	3,3	2,8
	2013	16 %	33 %	51 %	19 %	57 %	22 %	51 %	169 %	761 %	4,9	3,3	2,8
	2014	15 %	33 %	52 %	21 %	59 %	21 %	50 %	166 %	793 %	5,3	3,4	2,9
2015	15 %	33 %	53 %	22 %	59 %	20 %	50 %	165 %	802 %	5,4	3,4	3,0	
2016	15 %	31 %	54 %	23 %	60 %	20 %	48 %	159 %	814 %	5,4	3,5	3,2	
UAE	1998	12 %	33 %	54 %	18 %	63 %	16 %	42 %	200 %	849 %	7,1	4,2	2,6
	2009	11 %	33 %	56 %	16 %	65 %	18 %	36 %	219 %	799 %	6,1	5,0	2,4
Yemen	2006	14 %	37 %	50 %	16 %	59 %	15 %	51 %	201 %	658 %	7,2	3,4	2,3

3. CHARISMATIC LEADERS AND NATION-BUILDING

3.1 Introduction

“There are two Mustafa Kemal. One in flesh-and-bone who now stands before you and who will pass away. The other is you, all of you here who will go to the far corners of our land to spread the ideals which must be defended with your lives. I stand for the Nation’s dreams, and my life’s work is to make them come true.”

— Mustafa Kemal “Atatürk”, 1933

The construction of a national identity is fundamental to the development and consolidation of modern states (Tilly, 1993; Weber, 1976). It is also a highly relevant policy issue today. Indeed, many countries currently implement “nation-building policies” to overcome their ethnic, religious or linguistic cleavages (Alesina and Reich, 2015), as such cleavages are widely understood to impede cooperation, harm the quality of governance and heighten the risk of violence and institutional breakdown (Alesina et al., 1999; Fearon and Laitin, 2003; Habyarimana et al., 2009). Building a new national identity is however challenging, as culture tends to be quite persistent.¹ Besides, pre-existing local cultures might be inconsistent with the new national identity and top-down policies targeting them can amount to forced assimilation and therefore backlash or generate violence (Bisin et al., 2011, 2016; Dell and Querubin, 2017; Fouka, 2019).

A potential tool that can be used to quickly shape identity and drive cultural change are the actions of a national leader, as suggested by a large theoretical literature in economics.² This hypothesis, however, has not been tested to date. More generally, little is known about the channels through which a leader might contribute to the construction of a national identity.

This paper aims at filling this gap by answering the following questions: how are national identities constructed? Can individual leaders play any role, beyond policy choices and the reforms they implement? To address these questions, I focus on one particular

1. See Giuliano and Nunn (2021) for a recent literature review.

2. See for example Acemoglu and Jackson (2015); Akerlof and Holden (2016); Carvalho and Sacks (2021); Loeper et al. (2014); Verdier and Zénou (2018).

leader: Mustafa Kemal “Atatürk”, considered as the founder of modern Turkey and who is the main political figure behind the design and implementation of the Turkish nation-building reforms during the Turkish revolution (1923-1946), commonly called “Atatürk’s reforms.”

The Turkish historical context constitutes a unique setting to study the role of a leader in shaping identity. After the defeat of the Ottoman Empire during the First World War, a group of Ottoman soldiers led by Atatürk secured the Turkish territory against European invasion and implemented radical and authoritarian nation-building reforms to modernize and secularize the new nation-state (Zürcher, 2017). Beyond designing the new policies, Atatürk took a very personal role in supporting the new regime. From 1923, the first year of the creation of the Turkish state, until his death in 1938, he visited more than a hundred and fifty cities, i.e. a quarter of all Turkish cities, in order to rally citizens around his program.

I assemble a novel historical database with detailed information on the locations and dates of Atatürk’s visits from 1923 to 1938 at the district level. To study the impact of the visits on national identity, I exploit geographic and time variation in Atatürk’s visits in a difference-in-differences design. Moreover, I leverage the information available on the scheduling of his travels to exclude targeted visits and to focus only on districts visited along his routes. Using a large set of observable characteristics, I confirm that this group of incidentally treated districts is similar to nearby non-treated districts and therefore constitutes a plausibly quasi-random group of visited districts.

To measure local adoption of the national identity, I examine the first names chosen for newborns and whether or not they are in “Pure Turkish” (*Öztürkçe*), the new language introduced by the state as part of its homogenizing endeavor. Names constitute a particularly interesting outcome, widely used in the economics literature.³ Child nam-

3. See for example Abramitzky et al. (2016); Bazzi et al. (2020); Beck Knudsen (2019); Fouka (2019); Saavedra (2021), in the context of the U.S. and Scandinavia, and Ginsburgh and Weber (2020) for a recent literature review.

ing decisions are a major expression of cultural identity and more generally, language is a central determinant of national identity (Anderson, 1983; Weber, 1976). They can also be considered as a measure of a central state's capacity and ability to control its periphery (Scott et al., 2002).⁴ In the Turkish context, given the authoritarian and repressive nature of the regime especially toward ethno-religious minorities, the adoption a "Pure Turkish" name following a visit might reflect sincere adherence, opportunism or forced assimilation (Kuran, 1995). While I cannot precisely disentangle what giving a "Pure Turkish" name means, this variable however still constitutes a good proxy of the diffusion of the new national order and identity locally, which is likely to reflect a mix of the three interpretations.

To identify first names in "Pure Turkish" among newborns, I first digitize all booklets and newspapers published in the 1930s to disseminate the new words and created a list of common nouns in "Pure Turkish". Then, I use the universe of Turkish birth certificates between 1920 and 1950, a unique source in the Middle East, made available for the first time by the General Directorate of Population Affairs. Finally, I take advantage of the fact that first names in Turkey are common nouns and classify which first names are in "Pure Turkish". This allows me to create a granular and unique measure of national identity.

I find that Atatürk's visits generate a significant increase of 0.45 percentage points in the share of "Pure Turkish" names given to newborns, which represents an increase of 7 percent compared to the pre-visit mean. The magnitude of the effect grows over time, reaching almost 1.5 percentage points after fifteen years, which represents a medium-run increase of over 20 percent. It persists until twenty-five years after the visits before disappearing. The results are similar, and if anything larger, when using the restricted sample, which excludes targeted districts and focuses only on districts visited along the routes between two scheduled visit sites. They are also similar when using the estimator proposed by de Chaisemartin and D'Haultfoeuille (2020), which accounts for possibly

4. Indeed, as (Scott et al., 2002, p. 4) put it, "There is no State making without State naming".

heterogeneous and dynamic effects. I further show that these results are unlikely to be explained by selective migration, are robust to varying the sample definition and including only visited districts, and I also run a series of placebo tests supporting the validity of the empirical strategy.

Next, I examine the potential mechanisms behind the effect. First, to understand whether the visits laid the seeds for institutional changes locally, I collect new data on the locations and dates of creation of the “People’s Houses” (*Halk Evleri*), local cultural branches of Atatürk’s political party, the Republican’s People Party (*Cumhuriyet Halk Partisi*, henceforth CHP), established between 1932 and 1944. I find that the visits are strong predictors of the opening of a People’s House. Moreover, using the staggered establishment of the Houses, I show that they also have a positive effect on “Pure Turkish” names, which is stronger in places already visited. They therefore appear to serve as a complement to the individual actions of a leader. This is consistent with the fact that the effect of Atatürk’s visits on the adoption of “Pure Turkish” first names is significant only after a few years and is strongest after ten years. Rather than the visits themselves, it is the infrastructure and institution building they trigger that explains the diffusion of “Pure Turkish” names. Overall, this suggests that leaders and institutions act as a complements to construct a national identity ([Weber, 1921](#)).

Second, I examine two conflicting models of how a leader can contribute to the diffusion of a national identity: by rallying the masses, or by co-opting and persuading the elite. To test these competing views, I first collect additional information on the activities Atatürk held locally and I find that the effects are stronger when he met with local elites, and muted when he met only with the masses. I also collect new data sources to identify all cities with a former Ottoman nationalistic club, the “Turkish Hearths” (*Türk Ocakları*). The Turkish Hearths were elite-run associations created to promote Turkish nationalism, and which constituted a fertile ground for Atatürk’s ideology. I find that the effects are stronger in districts that had a Turkish Hearth. This suggests that Atatürk’s persuasion

effect is stronger in places already inclined to nationalistic values (Satyanath et al., 2017). Consistent with this effect being driven by the elite, I additionally find that the results are mostly driven by places with relatively high literacy rates. Finally, using biographical data of all members of the Turkish parliament between 1920 and 2010, I also show that the share of “Pure Turkish” names among deputies and their parents is systematically higher than the share in the overall population. These results confirm a large historiography concluding that Atatürk’s reforms were mostly successful among one segment of the population—the educated elite—which nevertheless enabled the regime’s survival (Atabaki and Zürcher, 2004; Lamprou, 2015; Tuna, 2018).

Overall, my findings suggest that an individual leader can contribute to the construction of a national identity, by influencing and co-opting an elite and by fostering institution building locally, who then complements the leader and influence identity at a broader level.

The main contribution of this paper is to build a very rich historical dataset for Turkey, with a novel measure of national identity. This paper also contributes to three main literatures. First, it adds to a growing literature on nation-building policies. This literature has focused on understanding why some states start implementing nation-building policies (Alesina et al., 2021; Alesina and Reich, 2015; Bandiera et al., 2019) and on analyzing several potential determinants of nation-building such as propaganda (Blouin and Mukand, 2019), military action (Dell and Querubin, 2017), education (Bazzi et al., 2018; Blanc and Kubo, 2021), population resettlement programs (Bazzi et al., 2019) or sport (Depetris-Chauvin et al., 2020). I focus on leadership, which this literature has not systematically examined with quantitative data yet and provide new evidence on the channels through a national identity is constructed. I also provide a novel measure of national identity, that can be measured at the local level and studied through time: the adoption of first names in the new language introduced by the state as part of its homogenizing endeavor.

This study also relates to the literature on leadership.⁵ On the theoretical side, this literature has emphasized the role of leaders in shaping various outcomes, including socio-cultural norms and identity (Acemoglu and Jackson, 2015; Verdier and Zénou, 2018).⁶ On the empirical side, there is now a well-established literature that shows that leaders matter for governance, state performance and in organizations (Bertrand and Schoar, 2003; Jones and Olken, 2005; Ottinger and Voigtländer, 2020). Recently, several papers have assumed a more micro perspective and studied the effect of individual leaders on various outcomes, such as the impact of the Pope on fertility (Bassi and Rasul, 2017), of the Forty-Eighters on social movements (Dippel and Hebllich, 2021), of Luther on the diffusion of the reformation (Becker et al., 2020) and of Pétain and Father Coughlin on votes (Cagé et al., 2020; Wang, 2021). My paper makes several contributions to this growing literature. First, this paper is the first to study a typical example of a charismatic leader—a military hero, seen as the savior of the people—on national identity. Second, I provide novel evidence on the channels through which a leader’s action matter and show that the effect is mostly due to his ability to rally an elite, and to implement local reforms, which in turn complement the leader’s efforts.

Finally, this paper contributes to the literature on the determinants and consequences of culture and identity. This literature has shown that identity and culture are major determinants of economic outcomes and behaviors (Akerlof and Kranton, 2000; Fernandez, 2010; Gorodnichenko and Roland, 2011). They tend to be persistent (Alesina et al., 2013; Giuliano and Nunn, 2021; Grosfeld et al., 2013; Grosfeld and Zhuravskaya, 2015; Spolaore and Wacziarg, 2013; Voigtländer and Voth, 2012) and jointly determined with institutions (Acemoglu and Robinson, 2021; Alesina and Giuliano, 2015; Bisin and Verdier, 2001, 2017). Little is known, however, about the short-term determinants of identity. I add to this literature by showing how a leader can quickly impact a national identity in

5. See Ahlquist and Levi (2011) for a general review of contributions from the political science, economics, and management literatures.

6. See also Akerlof and Holden (2016); Hermalin (1998); Murphy and Shleifer (2004); Loeper et al. (2014); Carvalho and Sacks (2021).

the short run and how the effect is mediated and strengthened by the establishment of complementary local institutions.

The paper proceeds as follows. Section 3.2 gives an overview of the historical background of the Turkish nation-building era and describes the language reform as well as Atatürk's visits. Section 3.3 presents the historical data on the visits, names and political institutions. Section 3.4 describes the empirical strategy. I present the main results of the impact of Atatürk's visits in Section 3.5. Section 3.6 explores the mechanisms. Section 3.7 concludes.

3.2 *Historical Background*

“The goal of our revolutionary measures is to bring the people of the Turkish nation to a modern and civilized stage.”

— Atatürk, 1925

This section outlines the history of the Turkish nation-building era, also called the “Turkish Revolution”, which started in 1923 and lasted until 1946. It presents the main institutional reforms implemented and describes the propaganda tools used by the state to spread the new identity, including Atatürk’s local visits.

3.2.1 *The Turkish National Revolution, 1923-1946*

After the defeat of the Ottoman Empire in World War I, a resistance movement led by Atatürk prevented European armies from invading today’s Turkish territory. Perceived as a hero and as the savior of what remained of the Empire, Atatürk became the first president of the new Turkish Republic in 1923. With his government, he implemented a set of radical and authoritarian top-down nation-building policies that affected all realms of society: political, legal, economic, social and cultural ([Zürcher, 2017](#)). Their goal was to create a nation-state that was “secular”, “modern”, “westernized” and built around the Turkish ethnicity, out of a six-century-old multi-religious, multi-ethnic and multi-lingual Empire.

Concretely, they abolished the Sultanate and the Caliphate and replaced the Sharia Law by European legal codes. They also implemented various nation-building policies used in other historical contexts ([Weber, 1976](#)): For example, they centralized the educational system and imposed a new and unique national curriculum built around secular values; they closed religious schools; they built railroads in order to connect the different regions of the new territory; they created a new national language. Atatürk and his government also passed a series of measures targeting the social and cultural life of Turkish citizens: They adopted the Western calendar and time, passed a law on clothing and

imposed the adoption of surnames, following the European patronymic system.

3.2.2 “Governing with words”: the Language Reform

“One of the significant characteristics of the nation is language. One, who regards himself as a member of the Turkish nation, should first of all and in every case, speak Turkish. If, someone, who does not speak Turkish, claims membership to Turkish culture and community, it would not be right to believe in this”.

— Atatürk, quoted in [Çagaptay \(2005\)](#)

Despite the Armenian genocide in 1915 and the Greek-Turkish population exchange in 1923, Turkey remained a highly heterogeneous society in 1923.⁷ The population included Muslim Turks, Jews and Christians but also non-Turkish Muslim groups and ethnicities such as the Kurds, Arabs or Lazes, among others.⁸ Each of these groups spoke their own dialects, which could have a different alphabet.⁹ Ottoman Turkish was the administrative language of the Empire. It was primarily a written language, based on the Arabic alphabet and known by a small educated elite. When the Republic was created, only 10 percent of the population was literate in Ottoman ([Pamuk, 2018](#)).

The Language reform was therefore a pillar of the Kemalist cultural revolution and remains one of its main legacies. Atatürk and his associates wanted to create a new Turkish language, easy to learn in order to increase the literacy rate, and which would be common across regions, religions, ethnicities and classes in order to foster a common national identity and assimilate minorities ([Türköz, 2018](#)).¹⁰

The reform was described by historians as a “catastrophic success” ([Lewis, 1999](#)), in that it was implemented quite quickly, in two main steps ([Aytürk, 2008](#); [Caymaz and Szurek,](#)

7. According to the 1914 census, Christians made up 20 percent of Turkey’s population, against merely 2 per cent in 1927 ([Çagaptay, 2004](#)).

8. Other ethno-religious groups include Muslim Georgians, Greek-speaking Muslims, Albanians, Macedonian Muslims, Pomaks, Serb Muslim, Bosnians, Tartars, Circassians and Abkhazes ([Çagaptay, 2004](#))

9. For example, Turks speaking Greek wrote Turkish in Greek characters. Armenians, Assyrians and Jews were speaking their own dialect or speaking Turkish, but with their own alphabets. Alevi (shia) Kurds were speaking the Zaza (Dimili) language. Atatürk, who was born in Thessaloniki (in Greece), was speaking a Turkish dialect specific from the Balkans.

10. “The [1934] Law will create a country speaking with one language, thinking in the same way and sharing the same sentiment”, Şükrü Kaya, Minister of Interior, 1934, quoted in [Ülker \(2008\)](#).

2007). First, the alphabet was latinized in 1928, in order to make Turkish more secular, as Arabic was seen as the language of Islam. The second step consisted in “purifying” the Ottoman vocabulary and was officially implemented in 1934.¹¹ The “purification” stage consisted in removing words of foreign origins (Arabic, Farsi etc.) and in replacing them with words in “Pure Turkish”, either invented or from the oral Turkish tradition and folklore, and which will constitute my main object of analysis.¹² This process was partly implemented in a top-down fashion, with scholars and linguists creating lists of words in Ottoman, with their synonyms in Pure Turkish.¹³ Citizens could participate voluntarily to the “word-collection mobilization” (*Söz Derleme Deferberliği*) (Lewis, 1999, p. 49) and send their own synonyms to the central state.¹⁴ As a result, a large number of booklets and dictionaries listing old ottoman words and their synonyms in the new language were published (Türköz, 2018). I collected and digitized all of these documents to create a comprehensive list of “Pure Turkish” words.

It is important to emphasize that the “Pure Turkish” language symbolizes in itself the new nation’s ideals, as conceived by Atatürk and his government (Aytürk, 2004; Mardin, 2002). The “purification” of the vocabulary underlines the homogenizing (and exclusive) aspect of the project, amounting to an ethnic cleansing of the language. As put by Atatürk himself in 1930: “*The Turkish nation which has proved its ability to defend its country and its full independence, should also free its language from the yoke of foreign language*” (Mango, 1999).¹⁵

11. For a description of the concrete implementation of the language reform and the institutions and actors which contributed to it, see Szurek (2013).

12. “*We will spare no effort in purifying our language of foreign rules and words, in making written language closer to spoken language, and in Turkifying the language usages of the state and of the sciences*”, General Regulations and Work Program of the Türk Dil Kurumu, adopted by the Fourth Congress of Turkish Linguistics], quoted in Szurek (2015).

13. “*Researchers were asked to travel the length and breadth of the country to record Turkish words which had survived only in provincial usage.*” (Mango, 1999, p. 495) .

14. “*To achieve [the Language Revolution] the first step is to collect Turkish language materials and create a dictionary. Each citizen should carry a brick for this blessed edifice. But the dictionary cannot be achieved without resorting to the memory of the nation.*” (Turkish Language Institute, quoted in (Türköz, 2018, p. 43).

15. This “linguistic nationalism” was accompanied by heavy propaganda. In particular, the effort was justified by a (false) theory, called the “Sun Language Theory”, according to which Turkish was the language from which all other “civilized” languages Kafadar (1995).

3.2.3 Spreading the words: Authoritarianism and State Propaganda

The central state used various propaganda tools to explain the ongoing reforms, especially the language reform. During the first years of the creation of the Republic, Atatürk himself led a massive campaign effort throughout the new territory, that is my main object of analysis and treatment. Later, Kemalist elites created local associations, the “Houses of People” (*Halk Evleri*), tightly linked to the single-ruling party, in order to propagate the reforms locally.¹⁶

Atatürk’s campaign: 1923-1938

Atatürk’s prestige as a hero of the Independence war (1919-1922) was used as an important propaganda instrument to gain support for the reform program (Zürcher, 2012). During the first years of the creation of the state, before most of the reforms were concretely implemented, Atatürk intensely travelled the territory and visited more than a quarter of all Turkish cities. The goal of the visits was to promote a unitary and national Turkish identity and to explain the new reforms program implemented by the central state. Atatürk was particularly keen on explaining and diffusing the “Pure Turkish” Language. The campaign effort also contributed to build his “charisma” and to establish a personality cult around his person. It also earned him the additional nicknames of “First Teacher” (*İlk Öğretmen*) and of name-giver, as he gave nicknames to some members of his government and renamed places during his visits.¹⁷ Figure 3.2 shows photographs taken during his visits and Appendix provides detailed examples of visits.

16. This section draws extensively on the fascinating historical work by Lamprou (2015) and Szurek (2013).

17. As described in (Türköz, 2018, p. 85), this nickname of the name-giver “begins with the performative act of naming of the new regime as *Cumhuriyet* (Republic). He then gives the surname *İnönü* to *İsmet Paşa*, for his bravery in the *İnönü Battlefield* in the War of Independence”. He also renamed geographic places and infrastructures during his visits (see his visit in Elaziz, for example, during which he renamed the city in Section ??).

The People's Houses: 1932-1944

"We have decided to raise the national unity and assemble all citizens under the roof of the People's Houses. The school is the classic institution a country has to prepare the nation for the future. However, in order to organize the modern nation as an entity, the usual methods and the regular efforts are not sufficient. If you want to become a nation in this century and form a national community, you will have to create the basis of a popular education in parallel and after the schools that will make the people work together as a unit."

— Recep Peker, CHP's secretary, 1932

The People's Houses were community centers opened and operated by the CHP between 1932 to 1944, with the duty to *"propagate the regime's ideology and policies to the population through the circulation, application and enactment of a variety of discourses and activities"* (Lamprou, 2015, p. 19). They were the ruling party's cultural branches. While the houses were officially presented as a non-political structures (Lamprou, 2015, p. 33), they did not possess any legal identity of their own, were tightly controlled by the CHP and exclusively financed by the state.¹⁸ All Houses were designed to have an identical structure and perform the same activities described in their by-laws (CHP 1932).¹⁹ Concretely, they organized various activities, such as sport events or theater plays to glorify the new republic. They diffused western literature and music while discrediting traditional *"alaturka"* music. They also organized adult classes to teach the new alphabet, convey positivist ideas, diminish superstitious beliefs and disseminate *"Western civilization"*. They organized collective listening of nationalist radio programs, broadcast in the new language (Ahiska, 2010). They also conducted research on local folklore to provide materials to the regime's historical and linguistic propaganda. In particular, they participated in the reform of the Turkish language through the collection of expressions in local use, *"ancient national fairy tales, sayings, proverbs and traditions"* (Lamprou, 2015),

18. They were created to replace the Turkish Hearths, former and independent Ottoman nationalistic associations, in order to control more closely civil society locally.

19. All Houses had nine branches: (1) Languages, History and Literature, (2) Fine Arts, (3) Theater, (4) Sports, (5) Social Assistance, (6) Courses, (7) Library and Publication, (8) Villages and (9) Museum and Exhibitions.

but also the organization of “Language party” (*Dil Bayramı*) every year on the 26th of September.²⁰

20. Appendix Figure C.13 displays a photograph taken during a “Language Party”, in Denizli, 1934.

3.3 *Data*

My baseline empirical work relates exposure to Atatürk during a visit between 1923 and 1938 to naming practices. In this section, I describe the data collected and used in my main empirical analysis.

3.3.1 *Exposure to Atatürk during a Visit*

To create the treatment variable, the visits made by Atatürk, I use a book by [Kocatürk \(1988\)](#), which compiles information on all Atatürk's official visits from various primary sources, including Atatürk's private diary, historical newspapers and additional archival records. For each of the 167 visited cities, there is information on the location, date, duration of the visit and the members of the visiting delegation. The average length of the first visit was one day and a half. I can identify cities Atatürk targeted and simple stops made on the way. For each visit, I systematically cross-reference the information from the book with other external sources, including Atatürk's biographies, academic articles, historical newspapers and municipalities' websites.²¹ Given that the birth certificates report only the district of birth of each citizen, all analyses which relate naming practices to visits are at the contemporary district level and not at the city level. I consider a district is visited when one of its cities is visited for the first time. In total, there are 154 visited districts out of the 973 contemporary districts.

I also look more precisely at the nature of the visits, and classify them depending on the activities Atatürk conducted locally.²² I distinguish whether Atatürk met with local elites and/or whether he met with the crowd. I consider Atatürk visited local elites if he had a meeting, lunch or dinner with local administrative, political or military personalities, or went with them in official institutions such as the municipality building,

21. See Section [C.1](#) for more detail on the sources, additional newspapers, and articles used. Table [A2](#) lists all visited cities, as well as the date of the first visit, the total number of visits and whether it was a visit targeted and planned by Atatürk. Table [A3](#) details for each itinerary the stops made on the way.

22. Out of the 154 visited district, I could collect information on the activities held for 122 visits.

the state house (*Hükümet Binası*) or military bases. I also code whether he pronounced a speech. Section ?? provides description of several visits to highlight how I classified the activities.

Figure 3.1 displays a map of the visited districts, as well as their timing. Figure C.12 provides more detail about the timeline of the visits. Interestingly, Atatürk conducted most visits in the first two years after the creation of the Republic, before the implementation of most of the reforms.

3.3.2 Naming Practices

Historical survey measures of cultural values usually do not exist. A good alternative, now widely used in the economics literature, is to look at naming practices.²³ A large psychological and sociological literature has indeed shown that first names signal attachment to a number of groups, such as age, ethnicity or religion (Coulmont, 2014; Mateos, 2013). Names are also visible, quite stable over time and often systematically collected by the state, which facilitates quantitative analysis. In the Turkish context especially, first names strongly signal social identity, ethnicity or religion (Bulliet, 1978; Gürpınar, 2012; Spencer, 1961).

I use child naming decision to measure citizens' reaction to the state ideology. To do so, I collected a unique data source: all historical birth certificates between 1920 and 1950.²⁴ The data contain information on the first names, dates and districts of birth of 15 millions of Turkish individuals born over the period. I describe below how I classify names to construct my main outcome variables.

23. See Abramitzky et al. (2016), Beck Knudsen (2019), Bazzi et al. (2020) or Fouka (2019) for some examples.

24. Source: General Directorate of Population Affairs.

“Pure Turkish” Names as a proxy for National Identity

I first identify newborns with first names in “Pure Turkish”, the new language introduced by the state as part of its homogenizing endeavor. To create this measure, I collected and digitized all historical booklets, dictionaries and newspapers published in the 1930s to spread the new language that are partly referenced by [Türköz \(2018, p. 115\)](#). This enables me to create a new and comprehensive list of more than 13,000 unique words in “Pure Turkish”.²⁵ Finally, I take advantage of the fact that in Turkish, first names are common nouns, and classify first names in the census data using this list of new words. I then compute the share of newborns with “Pure Turkish” names among all newborns, for all districts between 1920-1950. [Table 3.1](#) provides an overview of the ten most frequent first names in “Pure Turkish” and first names in Arabic or Turkish (traditional names), among children born in 1920 and 1940. Gulsum and Yasar are typically “Pure Turkish” names, while Mehmet and Fatma (respectively Turkish and Arabic) are typical traditional names. [Figure 3.3](#) displays the raw evolution of the share of names in “Pure Turkish”, for visited and non visited districts between 1920 and 1950. The peak observed in 1934, which coincides with the 1934 language reform. The figure also illustrates that that visited places adopted “Pure Turkish” first names more quickly.

Names in “Pure Turkish” reflect the diffusion of the new language introduced by the state locally. Given the authoritarian and repressive aspect of the regime, especially toward ethno-religious minorities, I however cannot disentangle precisely whether giving a “Pure Turkish” name is a sign of sincere adherence to Kemalism, opportunism or fear of repression ([Kuran, 1995](#)). I argue that they however provide, *a minima*, a good measure of the establishment of the new national order locally, and I discuss this in more detail in [Section 3.6.4](#).

25. [Figure C.3](#) displays examples of historical booklets listing new words and of a newspaper, publishing a dictionary with words in Ottoman and their synonyms in the new language.

Islamic and Arabic Names: Signaling Attachment to Tradition

Using Sakallı (2019)'s approach to identify religious male names, I classify as religious:

1. *Names with the suffix "of religion", i.e., names ending with "-(i)ddin" and "-(i)ttin", "-(e)ddin" and "-(e)ttin", "-(u)ddin" and "-(u)ttin"*
2. *Names names with the suffix "... of Allah" , i.e., names ending with "-(u)llah"*
3. *Names beginning with "Abd-", meaning "servant of ..."*
4. *The name of the Islamic prophet in its Arabic form, "Muhammad".*²⁶

I identify Arabic first names, using data from the Turkish Language Institute's website (*Türk Dil Kurumu*). Arabic first names were the most commonly used names before the creation of the Republic, and represent 75 percent of all first names in the birth certificates database before 1923. They were the main targets of the language reforms. It must be stressed many Arabic names can signal an attachment to Islam, and therefore that the two categories are not mutually exclusive.²⁷ I use Islamic and Arabic names to proxy for an individual attachment to Islam, to the former Ottoman Empire and/or to tradition.

Kurdish, Armenian and Jewish First Names: Signaling minority identity

Finally, I also create lists of Kurdish, Armenian and Jewish unique first names using the etymological dictionaries compiled by Sevan Nişanyan.²⁸

26. I also classify as religious other Arabic forms of "Muhammad, such as: Muhamed, Muhamet, Muhammed, Muhammet. The Turkish version of Muhammed, "Mehmet", is however not classified as religious.

27. Muhammed is an obvious example, but also Mahmud or Yusuf. The two categories are therefore not mutually exclusive.

28. See <http://turkadlar.com/>. The share of Kurdish and Armenian first names are strongly correlated with their population shares in the 1927 and 1914 censuses, as well as with the density of their villages.

3.3.3 *Additional data on nation-building tools*

I also collect disaggregated data on other nation-building policies used by Kemalist elites to propagate the new identity. I am able to track the railway expansion from 1925 to 1949, as well as the establishment of the People's Houses in cities and towns across Turkey from 1932 to 1944. Finally, I collect and digitize administrative school censuses tracking the number of schools, teachers and students at the district level, for 1925 and between 1932 and 1945.

3.3.4 *Historical and geographic covariates*

Using various primary and secondary sources, described in Appendix C.1, I obtain a rich set of historical covariates before 1923 at the district level, with information on local development, culture and politics: literacy rates, road network in 1928, number of cities and their administrative status (province or district), ancient trade roads, density of minority villages, distance to Ankara and Istanbul and distance to former Ottoman nationalistic associations, the "Turkish Hearths" (*Türk Ocakları*), created in 1912, that I use to measure pre-treatment adherence to Turkish nationalism.²⁹ I also collect geographic information: elevation, coastal cities, mean of annual precipitation and temperature and indexes for various crop suitability from the GAEZ database.

29. The clubs were created by local elites and intellectuals, influenced by Ottoman nationalist thinkers such as Ziya Gökalp.

3.4 *Empirical Framework*

In this section, I start by describing the main determinants of the visits in order to document the strategy behind his campaign effort. I show that Atatürk visited the most populated districts, among the most likely to adhere to the new order.

Then I present my empirical strategy, which is twofold. First, I implement a difference-in-differences strategy, using time and geographic variation in Atatürk's visits to estimate his impact on naming practices. Second, I take advantage of the quasi-random inclusion of districts in the campaign due to their location along the road connecting the districts that were the target. I estimate specifications in which all districts that may plausibly have played a role in determining the itinerary of the campaign are removed from the sample (henceforth the "targeted" or final destinations), and which include only districts crossed during his travel. Using this group of incidentally treated districts, I find that the estimated effects are virtually identical to that using the full sample and that in both cases, there were no differential pre-trends in the outcome variable between visited and non visited cases.

3.4.1 *Understanding the Logic of the Visits*

In a first step, I model the probability of a visit to a district as a function of the historical and geographic covariates collected in order to shed light on the strategy behind Atatürk's campaign effort.³⁰ Table 3.2 reports probit estimates of the probability of being visited, for all visits (Columns 1 and 2) and for visits by year (Columns 3 to 10). As shown in Column (4) in Table A4, development indicators (population size, administrative statuses, city density) strongly predict the visits, as well as the distance to a former nationalistic club. Atatürk also went to places with less minority villages, although this is not always the case across years. It is interesting to note that the only predictor that

³⁰. I was not able to find official document describing the strategy chosen for the visits. I therefore use the data collected to document it.

is systematically significant, no matter the timing of the visits, is the population size. Overall, the campaign effort was targeted towards large and developed localities, more Turkish, and closer to the new capital Ankara.

3.4.2 *Sample Restriction*

In order to identify causal estimate, I additionally perform the following sample restriction. I identify and remove the starting and ending points of all itineraries. The historical sources, unfortunately, do not provide additional information about why they were selected nor describe the organization of the trajectory. I simply observe the final destination, as well as the stops made along the route. Table A3 lists all itineraries and defines for each of them the start and end points as well and the stops Atatürk made. I further restrict the sample to districts lying on his way and crossed during his travel.³¹ This procedure restricts my sample to a set of 272 districts over the 973, among which 92 were visited, displayed in Figure 3.4. These visited districts were plausibly quasi-randomly included in the campaign due to their lying along the road connecting the cities that were the true target. Figure 3.5 plots the standardized beta coefficients of bi-variate regressions of the variables listed on the left on the visit status, with province fixed effects and robust standard errors in the full sample (Panel A) and in the restricted sample (Panel B). Visited and non-visited districted in the restricted sample appear to be very similar and balanced over the set of historical and geographic covariates, and constitute a plausible comparison group to visited districts on Atatürk's way. There are no differences in the main predictors of the visits: population size, administrative status, connectedness and closeness to a former Ottoman nationalistic club. The remaining covariates significantly different between both groups are the number of Armenian and Arab villages, the number of minority schools and religious buildings and whether the

³¹. I identify them using a GIS software. I compute the least cost path for all itineraries during which Atatürk made a stop, using new data on the road network in 1928, on railway networks between 1923 and 1938 as well as information on topography.

districts are on the coast. I will systematically control for these characteristics. It should be noted that my identification strategy lies on the fact that there are no differential pre-trends between the two groups—and not that they are similar in terms of observable characteristics. Finally, Table A4 presents summary statistics of the main variables in the database, depending on the visit status, in the two samples. My specifications run on the full sample will systematically include as covariates the characteristics that are significantly different between the treated and control group, that is for the (log.) of the total number of birth, whether or not the district has a province or district center (in 1935), the (log.) of the distance to the nearest railway in 1919, road in 1928 and former major trade roads, the total city density, the literacy rate, the number of Kurdish villages, the (log.) of the distance to the nearest Ottoman nationalistic club, to Istanbul, to the border, an indicator on whether or not the locality was occupied during the Independence War (1919-1922), the density of minority schools and religious buildings, as well as a set of geographic covariates (whether or not the district is on the coast, the average temperature, elevation and suitability indexes for cultivated crops).

3.4.3 Main Specifications

Difference-in-Differences at the district level

I estimate the effect of the visits on naming practices using a difference-in-differences model with district and year fixed effects between 1920 and 1950, given by the following specification:

$$Y_{dt} = \beta(Visited_d \times Post_{dt}) + \theta X'_d * \gamma_t + \delta \text{Log}(NbBirth_{dt}) + \alpha_d + \gamma_t + \epsilon_{dt} \quad (3.1)$$

where d and t index districts and years respectively. Y is the share of newborns with a “Pure Turkish” first name. My main treatment $Visited_d \times Post_{dt}$ is a variable equal to one when a district is visited for the first time and which stays equal to one the following

years. α_d and γ_t are districts and year fixed effects, which allows me to control for any time-invariant differences between districts and for year-specific shocks common to all districts. X is a vector of pre-treatment time invariant historical and geographic controls, interacted with year dummies. Only the total number of births is time varying, and is also included as control. The coefficient β on $Visited \times Post$ is the coefficient of interest and captures the additional change in the share of newborns with Pure Turkish names following a visit, in percentage point. I cluster all errors at the district level, as it is the level of treatment. $\epsilon_{d,t}$ is an error term.

In the baseline specification, I include as covariates all characteristics significantly different between the treated and control groups, as displayed in Column (4) of Table A4. The main identifying assumption of this strategy is that, in the absence of a visit, the average change in the share of name in Pure Turkish in treated and control districts would have been the same—the two types of districts would have continued to experience parallel trends.

Event-Study Specification

To test this assumption, I check for potential pre-trends by estimating the following event-study specification:

$$Y_{dt} = \sum_{k=\underline{C}}^{\bar{C}} \beta_k E_{dt}^k + \theta X'_d * \gamma_t + \delta \text{Log}(NbBirth_{dt}) + \alpha_d + \gamma_t + \epsilon_{dt} \quad (3.2)$$

where the event-time dummies E_{dt}^k are defined as: $E_{dt}^k = \mathbb{1}[t = \tau_d + k] \forall k \in (\underline{C}, \bar{C})$, $E_{dt}^{\bar{C}} = \mathbb{1}[t \geq \tau_d + \bar{C}]$, and $E_{dt}^{\underline{C}} = \mathbb{1}[t \leq \tau_d + \underline{C}]$, where $\mathbb{1}[\cdot]$ is the indicator function and τ_d is the first year when a district is visited by Atatürk. I normalize $\beta_{-1} = 0$ and set $\underline{C} = -5$ and $\bar{C} = +27$. X is the same vector of time invariant controls, interacted with year dummies. The β_k coefficients give the full path of dynamic effects, that is pre-event effects necessary to check for pre-trends, but also post-event effects. I additionally use the estimator pro-

posed by [de Chaisemartin and D'Haultfoeuille \(2020\)](#), which is robust to heterogeneous and to dynamic effects.

3.5 Main Results

3.5.1 Effects of Atatürk's Visits on Naming practices

Pure Turkish and Arabic Names

Table 3.3 presents the main results on the effect of Atatürk's visits on the the share of "Pure Turkish" first names among newborns and of other types of names, such as Arabic, Religious or minority names. Results are displayed for the full sample (Panel A) and for the restricted sample (Panel B), excluding targeted districts. As shown in Column (1), visited districts are more likely to adopt first names in the new language introduced by the state to homogenize the population, in both samples. A visit is associated with an increase in the share of Pure Turkish names of 0.45 percentage point, which represents a change of around 7 percent compared to the sample mean before the visits. The estimated coefficient is even larger in the restricted sample: a visit is associated with an increase of 0.56 percentage point in the share of newborns with a "Pure Turkish" first name, which represents an increase of 10 percent compared to the sample mean pre-treatment.

Column (2) shows that Atatürk's visits are associated with a decrease in the share of Arabic first names, which is consistent with the fact that "Pure Turkish" words were precisely introduced to replace words in Arabic in the Ottoman language. Arabic first names were also the most commonly used, by Arabs but also by Turks or Kurds. Their use therefore reflects tradition more broadly and not a signal of attachment to the Arab minority.

Figure 3.6 presents the results of the event-study specification. It plots the coefficient estimates $\{\beta_k\}_5^{15}$ given in equation 3.2. Prior to his first visit, the estimated difference between treated and control districts is statistically indistinguishable from zero, in both samples. The F-stat for the joint significance of the pre-reform estimates on "Pure Turkish" names equals 0.96 (p-value 0.41), thus confirming the absence of pre-trends and providing support for the parallel trend assumption. Following the visits, the share of

“Pure Turkish” first names in visited districts increases significantly relative to the share in control districts. The effect persists and its magnitude is growing over time, reaching 1.5 percentage points after fifteen years, which represents a medium-run increase of over 20 percent. Then, the effect decreases and disappears after twenty five years. Conversely, the share of Arabic first names significantly and persistently decreases following a visit, in both samples as well.

In order to better understand the dynamic of the effect, I also use a difference-in-differences specification of the form given in equation 3.1, but using as main treatment an interaction between a dummy indicating whether a district has ever been visited and a variable which indicates the numbers of years since a visit. Results are displayed in Table 3.4 and measure the average effect of a visit on names, for any year following a visit. On average, each additional year following a visit significantly increases the share of “Pure Turkish” name by 0.05 percentage points in the restricted sample. Finally, to understand which years drive this average effect of an additional year, I use a piecewise linear regression model. Results are displayed in Table A5 and show that Atatürk’s effect on “Pure Turkish” names appears after five years and is strongest between 10 and 15 years after a visit in both sample (Column 1).

A number of recent studies show that, in the presence of heterogeneous and dynamic treatment effects, the coefficients on the leads and lags of the treatment variable in an event study might place negative weights on the average treatment effects for certain groups and periods (Borusyak et al., 2021; Callaway and Sant’Anna, 2020; de Chaisemartin and D’Haultfoeuille, 2020). This might be particularly true in my setting, where the treatment effect is likely to be heterogeneous and to vary depending on the type of activities Atatürk held locally, the time he spent there or local characteristics. To address this concern, I use the estimator proposed by de Chaisemartin and D’Haultfoeuille (2020).³² The results are presented in Figure 3.7 (a): Similarly to my baseline event

32. I use the software packages `did_multplegt` developed by de Chaisemartin and D’Haultfoeuille (2020).

study, these results indicate that first names in “Pure Turkish” increases following a visit by Atatürk, whereas before the visit, the effects are not distinguishable from zero.

Overall, Atatürk’s visits have a positive effect on “Pure Turkish” names: the effect takes some time to appear, then increase until fifteen years of a visit, before decreasing and disappearing.

Islamic and Minority names

Given that secularism was a pillar of the Kemalism, I look the impact of Atatürk’s visits on religious names. Results are displayed in Columns (3) and (4) of Table 3.3. Atatürk’s visits are associated with a decrease in religious first names, but the estimate is not statistically significant (Column 3) , except when looking at the name Muhammed only (Column 4). This result is however not robust to using the restricted sample.

A possible interpretation for this less precise effect is that Atatürk’s propaganda did not manage to deeply secularize people as it did not significantly impact the choice of Muslim first names. This result might reflect an important and well-known paradox of the Turkish nation-building and secularizing reforms: The new identity had to be “purely Turkish” and, given that most Turks were Muslims, promoting the Turkish identity also promoted a Muslim identity. In other words, being Muslim was an implicit condition to belong to the new secular Turkish nation, at the expense of other religious and ethnic identities (Çagaptay, 2005; Fabbe, 2019). This result should nevertheless be interpreted with caution: the classification of religious first names focuses only on males, and is not exhaustive. In particular, it does not include many Arabic first names, present in Column (2) that could be used to signal an attachment to Islam as well.

As shown in Column (2) of Table 3.4, on average, each additional year following a visit significantly decreases the share of Arabic name by 0.08 percentage points in the restricted sample. I also examine the dynamic of Atatürk’s effect on Arabic names in Table A5. In both sample, the effect is negative and significant for the five first years.

Finally, results are similar when using the estimator [de Chaisemartin and D’Haultfoeuille \(2020\)](#)’s estimator, as shown in [Figure 3.7 \(b\)](#).

Finally, as displayed in Columns (5) to (7), Atatürk had no effect on Armenian, Kurdish and Jewish first names. Minorities did not give up their names. [Figure C.14](#) displays the coefficients estimated by the event-study specification and confirms the absence of effect.

Magnitude of the effect and persuasion rates

In order to quantify the magnitude of the effect of a visit and compare it to other studies, I compute persuasion rates following [DellaVigna and Kaplan \(2007\)](#) and a large number of studies in the empirical persuasion literature ([Adena et al., 2015](#); [Cantoni et al., 2017](#); [DellaVigna et al., 2014](#); [DellaVigna and Gentzkow, 2010](#); [Enikolopov et al., 2011](#))

In my context, the persuasion rate is the estimated percentage of individuals (parents or future parents) who did not initially have the behavior the visits aimed to propagate (using the new language to name children) but who decided to adopt it as a result a visit by Atatürk.

Formally, the persuasion rate is given by: $f = 100 \times \frac{y_t - y_c}{e_t - e_c} \frac{1}{1 - y_0}$, where $y_t - y_c$ is the naive treatment effect, a simple difference of outcome between the treated and control groups; $e_t - e_c$ is the difference of exposed individuals between the two groups. $\frac{1}{1 - y_0}$ represents the fraction of the population with the intended behavior. To compute the persuasion rate, I follow [Cantoni et al. \(2017\)](#) and estimate the fraction of individuals who would have the desired behavior in the absence of a visit. To do so, I predict naming practices using my baseline regression model. For individuals living in visited district, I subtract the treatment effect of a visit. I then average the predicted outcomes for those who live in visited districts and those who live in non-visited districts and use this to calculate the fraction of the sample who would not have the desired behavior

in the absence of a visit. I then use this share to compute the persuasion rate.³³ I find a persuasion rate of 9.5 percent, which ranges in the middle of the rates found in the literature (typically varying between 6 and 20 percent).³⁴ It is slightly smaller than the other persuasion rates estimated in the literature that focus on direct face-to-face contact (15 percent) found in [Gerber and Green \(2000\)](#). It is also smaller than the 28 percent persuasion rates found in [Wang \(2021\)](#) for Father Coughlin, another charismatic individual but who used mass radio to disseminate his message.

3.5.2 *Addressing Identification Challenges*

In this section, I probe the robustness of the baseline results to various potential threats to identification.

Alternative Samples and Sensitivity Analysis

Identification of my main coefficients hinges on the assumption that districts yet to be visited and non-visited districts form a credible counterfactual for visited districts, after accounting for time-invariant (observed and unobserved) differences between districts and year-specific shocks common to all districts. As shown in [Table A4](#), visited and non-visited districts are unbalanced along several observable characteristics, mostly population size and connectedness described in [section 3.4](#). To show that this imbalance does not drive my results, I follow [Hainmueller \(2012\)](#) and use entropy balancing to re-weight observations so that visited and non-visited districts have the same mean and variance for all historical and geographic covariates after subtracting district and year fixed effects. [Table A7](#) shows that re-weighting following this procedure leads to a bal-

33. Given that the outcome variable is non-binary, I calculate the persuasion rate based on a transformed dependent variable, which equals one if the outcome is greater than or equal to the median outcome, following [Cantoni et al. \(2017\)](#).

34. [DellaVigna and Kaplan \(2007\)](#) find a persuasion rate from Fox News of approximately 3-8 percent, and [DellaVigna et al. \(2014\)](#) find a persuasion rate of 4-5 percent for Serbian radio in Croatia. [Enikolopov et al. \(2011\)](#) find an 8 percent persuasion rate in an analysis of an independent Russian television station's effect on voting for the opposition parties it supported.

anced sample: there is no significant relationship between the treatment dummy and any of the district level pre-treatment characteristics. As displayed in Figure C.15, the results are very similar, suggesting that misbalance does not drive our results. Finally, I also verify that the results are not driven by influential observations and robust to excluding one district at a time from the sample. Results are displayed in Figure C.20.

One might still be concerned that never-visited districts do not provide a suitable counterfactual for visited ones. To test whether my main estimates are explained by the contrast to never-visited or by the staggered timing of visits, I further restrict the sample to visited districts only, and replicate the analysis. This allows me to compare the outcomes of visited district in event year k to the outcomes of future visited district. As shown in Figure C.16, I find very similar estimates, which points to the event as the primary driver of my estimated effects.

Accounting for Selective migration

Another key threat to the difference-in-differences identification strategy is endogenous sorting across districts. Individuals more likely to adopt “Pure Turkish” first names—and to follow the new order—could be migrating into districts in response to or in anticipation of a visit or of subsequent changes in the visited locality. Similarly, targeted and repressed minorities might be leaving these localities. The effect would then be driven by this compositional change and would be most likely biased upward. This is an important concern in this setting, given that large population movements were happening at the time. After the Greco-Turkish war of 1919-1922, 1.2 million Greek Orthodox were forcibly resettled from Turkey to Greece and inversely Muslim Greek resettled in the new Turkish territory in 1923. To account for this, I focus only on individuals with fathers born in the same district, given that the historical birth certificates provide information on the fathers’ places of birth.³⁵ Results are very similar, as displayed in Figure

35. If the father is born in another district or if the father’s place of birth is missing, I do not take into account the observation to compute my outcome variables.

C.17. Coefficients estimates are of similar magnitude, as shown in Table A6.

Placebo exercises

I also conduct two placebo exercises. First, I fix the total number of districts receiving the treatment, and randomly draw the districts which receive the visits and the year of the visit. Second, in order to confirm that the visits are not associated with a concurrent increase in “Pure Turkish” names in other nearby districts from the same province in the same year, I randomly draw placebo-event districts among districts that did not have a visit from the province-years in which other districts had a visit. I repeat both exercises 500 times, comparing the distributions of the point estimates and their t-statistics for the effect of such placebo treatments with those for the actual treatment. Results are presented in Figure C.19. In both cases, the baseline coefficient and its t-statistics from the estimation of the effect of the true event are outside of the corresponding distributions for the placebo events. Event study treatments are not associated with a concurrent increase in Pure Turkish names in other districts of the same province in the same year of a visit.

Discussion on the potential confounding effect of concurrent policies

Another threat to the identification assumption is the potential confounding effect of other policies and legislations happening concurrently, that is, if the visits happen in districts that are systematically subject to additional—and similar—nation-building policies or events at the same time, independently from the visits. The effect observed would therefore not be due to the visits themselves. This is however unlikely to be the case. First, the lack of pre-trends and the robustness of the results in different restricted samples, including only visited districts, already provided reassurance that visited places were not on a different political trajectory.

The other policy interventions used by Kemalist elites to homogenize the territory and diffuse the new ideology were media, school constructions, railway expansion, local

associations—the People’s Houses— and similar visits made by the Prime Minister İsmet İnönü. I investigate in more detail in Section 3.6 the role of the railway, of the Houses and of the Prime minister’s visits, and how they interact with Atatürk’s visits, and show that they do not fully drive the results, and act as positive complement to Atatürk’s visits.

Media are also unlikely to play a major role: mass radio was established as a propaganda tool in 1938, after the last visit. Given the low literacy rate (10 percent), propaganda via newspapers is also unlikely to fully explain the results. This is also the case for school construction. As shown in Figure C.18, most to the school expansion also happened at the end of the 1930s and mostly in the 1940s. Similarly, the Houses were established between 1932 and 1944.

Railway expansion, however, happened between 1925 and 1949, at the same time as the visits. Railroads could play a role similar to the one of Atatürk and convey information or could make it easier for the central government to control the locality (Cermeño et al., 2021; Voigtlaender and Voth, 2014; Weber, 1976). As shown in Table A12, railway seems to have a positive impact on “Pure Turkish” names, that is smaller in magnitude than the effect of a visit. As shown in Column 2, however, when controlling by the railway expansion, the effect of a visit does not disappear. There is however, weak evidence that the railway and the visits acted as complement, as shown in Column 3, that estimates the heterogeneity of the effect of the railway depending on whether or not the district has already been visited

3.6 Mechanisms

Having documented that Atatürk's visits had a local impact on national identity as measured by naming practices, and that this impact takes time to appear and is strongest after ten years, I now explore the channels that might explain his persuasiveness.

3.6.1 *Leader and institutions: Complement or Substitute?*

First, I investigate whether the visits had any impact on the local institutional landscape. To do so, I focus on one institution, the "People's Houses", which were community centers opened and operated by the Kemalist party between 1932 to 1951. More precisely, the Kemalist regime closed the "Turkish Hearths", created during the Ottoman Era, from grassroots initiatives, and created the Houses in order to control civil society more tightly, and to provide a unified propaganda throughout the territory. The Houses had a similar propaganda role as Atatürk visits. Additionally, they organized various activities aimed precisely at propagating the new language, such as adult classes, workshops to collect new words, collective listening of the radio program starting from 1938 — and broadcasted in the new language—or celebration of the language every 26th of September (Lamprou, 2015).

Atatürk's Visits predict the opening of the People's Houses

In a first step, I investigate whether the visits predict the opening of the People's Houses. I create a geo-coded database at the town level (35,000 villages and cities, with data on historical population for 21,000 of them) and create a similar set of historical and geographic covariates as in the main district-level database, using QGIS software. I also compute for each town the distance to the nearest visited city and to the nearest House, using new archival records on localization of the Houses and their year of creation described in section C.1.

I find that the visits are consistent predictors of the opening of a People’s Houses in the 1930s, as displayed in Table 3.5. Column (1) displays the results of an OLS model regressing the distance to the nearest visit on the distance to the nearest House, with all historical and geographic covariates collected and district fixed effects. Column (2) shows the results of a similar model, including only as covariates predictors selected using a Lasso procedure, which selected the distance to a visit. Column (3) estimates a probit model, with a binary variable “has a House” or not as main outcome.³⁶ The visits are strong predictors of the opening of a House.

Table A10 present results from similar regressions, but distinguishing between the Houses which opened between 1932 and 1938 and on those which opened later. The visits best predict the opening of the first group of Houses. Figure 3.13 shows binscatter plots of the corresponding relationship, with controls selected by LASSO and district fixed effects. Results are robust to correcting for spatial Correlation following Colella et al. (2020); Conley (1999); Hsiang (2010); Kelly (2019), as shown in Table A11.

I however cannot disentangle whether this is due to additional resources sent to these localities, or to local elites being more zealous and implementing more quickly the reforms locally. It however shows that the visits is linked to a change in the local institutional landscape.

Impact of other nation-building tools on naming practices

To examine whether the Houses and the leader are complement or substitute, I investigate the effect of the establishment of the Houses on local naming practices. I implement an event-study of the form given by equation 3.2, where my outcome variable is the share of “Pure Turkish” first names but where my treatment variable is a variable which equals 1 the first year a House is opened in a given district, and which stays equal to one.

36. A city is considered as having a House if there is a House within 5 kilometers.

Results are displayed in Table 3.6, Panel A. The establishment of a House is associated with an increase in Pure Turkish name of 0.55 percentage points (Column 1) and a decrease in Arabic names of 0.60 percentage points (Column 4). There are no clear effects on religious names (Column 10) as it was the case for the visits. As shown in Column (3), the effects are stronger in districts that have been visited, suggesting a complementarity between institutional propaganda and the campaign effort made by the leader.

However, the Houses are also associated with an increase in Kurdish names of 0.09 percentage points (Column 7), which represents a change of 8 percent compared to the mean pre-treatment. This effect can be interpreted as a cultural backlash to an institution and propaganda effort that targeted the local's population identity. This effect is consistent with a large literature that has shown that identity may be strengthened in the face of policies aimed at integration both theoretically (Bisin et al., 2011, 2016; Carvalho, 2013) and empirically Fouka (2019); Sakallı (2019). According to these studies, families which perceive their cultural traits as being in threat of extinction inculcate even more those traits to their children and ensure persistence. It also confirms a large historiography that has emphasized the various ways localities have negotiated and with the central state and developed day-to-day forms of resistance against the "Turkification" policies, in particular among minorities and Kurdish areas (Aslan, 2011; Yilmaz, 2013).³⁷

This result is consistent with recent work by Caesmann et al. (2021), who find that propaganda can persuade and generate a backlash, in the context of the 1932 Nazi marches in Hamburg. In my context, the backlash is visible only when the propaganda institutionalizes—and not when it was only under the form of the visits. It should however be noted that Atatürk did not go much to South-Eastern Kurdish regions, as shown in Figure 3.1, which might also explain this absence of backlash.

Figure 3.14 plots the coefficient estimates of the event study of the effect of the opening

37. As Lamprou (2015), in non-Turkish areas, including Kurdish South-Eastern regions, to be "turkified", the Houses were "isolated state colonies in the middle of the vast ethnic linguistic and cultural otherness they were supposed to eradicate" (p73).

of a People's House on (a) Pure Turkish names and (c) Kurdish names, confirming the absence of pre-trends. Following the opening of a House locally, the share of first names in Pure Turkish rises significantly relative to the share in control districts the first two years, but then decreases and is statistically indistinguishable from zero. The share of Kurdish first names among newborns, however, increases persistently. The Figure also displays graphically the heterogeneous effects depending on whether or not the districts was visited before. As we can see in Figure 3.14 (b) the positive effect on "Pure Turkish" names is completely driven by districts that were also visited by Atatürk. Reversely, the effect on Kurdish names is muted in places that were already visited (Figure 3.14, d).

Overall, these results suggest that the visits led to the establishment of institutions locally, which in turn, also has a positive effect on first names, and acts as a complement to the leader. This partly explains why the effect of Atatürk's visits is strongest after ten years. Indeed, as shown in Figure ??, Houses are established on average ten years after a visit.

3.6.2 *Co-optation of local elites*

A central conclusion in the historical literature on Kemalism is that the nation-building reforms were mostly successful among a segment of the population: the educated and urban upper-middle class and the elite (Atabaki and Zürcher, 2004). In this section, I quantitatively investigate this hypothesis.

Heterogeneity Analysis with respect to the activities conducted locally

In order to understand the role of the elite, I start by analyzing the heterogeneity of the effect of Atatürk's visits with respect to the activities he conducted locally. To do so, I collect detailed information on the activities held for 122 of his 154 visits, and classify them depending on whether or not he met with local elites, the masses and/or made a speech, using using historical newspapers, academic articles and on contemporary

municipalities and districts' websites, which often have a section on Atatürk's visit if he went there.³⁸

Results are displayed in Table 3.7, for the full sample (Panel A) and for the restricted sample, excluding targeted districts (Panel B). While the differential effect of the activities held is not clear in the full sample, visits where he met with local elites drive most of the effect in the restricted sample, providing suggestive evidence that co-optation of elite played a role, especially among non targeted districts. Figure 3.9 plots the corresponding dynamic coefficients, showing that there are no pre-trends—places where he met with elites were not on a different political path before the visits. Visits where he met with the masses also has an effect, that is however smaller in magnitude, as shown in Column (3).

Heterogeneity Analysis with respect to the distance from a former Ottoman nationalistic club

To further examine the role of local elites, I explore the heterogeneity of his impact depending on the distance from former nationalistic associations, the “Turkish Hearths”, which were created by an urbanized elite and intellectuals in 1912, influenced by Ottoman nationalist thinkers such as Ziya Gökalp and close to the Committee of Union and Progress (CUP) and Young Turk movement. Their goal was to spread Turkish nationalism locally.³⁹ I collect new archival data to localize the Hearths and use the distance to a Hearth as a measure of Turkish nationalism strength locally, before the first visit. Results are displayed in Table A9, Column (5). The effect of the visits are larger in districts closer from a Hearth, that is in places that constituted a more fertile ground to the Kemalist propaganda, mostly in the restricted sample. Figure 3.10, which displays the coefficients from an event-study, using a binary variable to capture the distance from a former Hearth

38. Among the 122 visits for which information is available, he met local elites during 32 visits, the crowd during 39 visits and both the elites and the crowd during 51 visits. He made a speech 30 times in total. Section C.1 in Appendix provides detailed description of several visits and examples of sources to illustrate how I classified the visits.

39. In 1931, they were however closed by the Kemalist regime, and replaced by the People's Houses, that were directly linked to the central state and more tightly controlled compared to the Hearths.

as heterogeneity variable.⁴⁰ As we can see, there are no pre-trends among places closer from a Hearth. Appendix Section ?? gives an example of a visit Atatürk made to Aydın during which he precisely asked members from the Turkish Hearth to help him in his endeavor.⁴¹

Heterogeneity Analysis depending on Literacy Rates

I also study the heterogeneity of the effect of the visits depending on the literacy rates, computed using the 1927 census. I construct a dummy variable which equals 1 if the district has a literacy rates above the median, and 0 if if not. Figure 3.11 and shows that the effect is driven by places with a high literacy rate, adding evidence that the results are mostly driven by the elites.

Evolution of Pure Turkish names among the Elites

Finally, to provide additional suggestive evidence that it was mainly the elite that followed the new order, I digitize the biographies of all Turkish deputy members between 1920 and 2010, from the Library of the Grand National Assembly of Turkey (*Türkiye Büyük Millet Meclisi*).⁴² The books contain information on the first names and dates of birth of 6,022 deputies, born between 1844 and 1977, as well as the first names of their parents.⁴³ I use this source to compute the share of Pure Turkish names among deputies and their parents over time and to compare it to the overall population. As shown in Figure 3.12, deputy members and their parents display a systematically higher share of Pure Turkish names compared to the overall population.

Overall, these additional results suggest that the effect is driven by a specific segment

40. The indicator is equal to 1 when the district is within 30 km of the former Hearth.

41. This result is consistent with existing work emphasizing the role of social capital in promoting ideologies, for better or worse (Satyanath et al., 2017).

42. The books, in four volumes, are available in pdf format in the following [website](#).

43. I assume that on average, deputy members' fathers had their children at 25 years old, and their mothers' at 20 years old. Results are unaffected depending on the assumption made on the date of birth of the deputy members' parents.

of the population, the elite. The People's House act as an endogenous intermediating variables that strengthen the effect. I cannot show whether the establishment a House is due to the fact that these localities receive more funding following a visit or to the fact that the elite actively participated in their formation. Nevertheless, Atatürk's effect appears to be due to his ability to co-opt an elite, then adopts itself the new national identity and contributes to the implementation of the reforms locally.

3.6.3 *An idiosyncratic effect? Comparing Atatürk's and İnönü's visits*

Is the effect due to something specific about Atatürk, that could be his charisma? In order to see whether the main results are due to something specific about Atatürk, I compare his effect to the effect of his Prime Minister and second man: İsmet İnönü. İnönü was a central political figure during the nation-building era and became President after Atatürk's death. Anecdotal evidence suggests that he was less charismatic than Atatürk (Metin, 1998).⁴⁴ He made a similar campaign effort throughout Turkey, in order to convey Kemalist ideas to the periphery.

I collect similar data on İnönü's visits: In total, 294 districts are visited, 49 by Atatürk only and 140 by İnönü only and 105 districts were visited by both men, together or not, as described in Appendix Table A1.⁴⁵ I use this variation to test whether Atatürk had an idiosyncratic effect. If Atatürk's effect was due to something specific about him, one would expect him to have a larger influence on naming practices in visited districts. I use difference-in-differences models of the form given by equation 3.1, exploiting time and geographic variations in both leaders' visits. I use two different samples: the full sample and a sample which excludes the province centers, which are the most populated

44. In his book on İnönü's career and life, Heper emphasizes İnönü's large influence on Turkish politics and challenges the most commonly held view according to which İnönü was only Atatürk's "second fiddle". He nevertheless "acknowledges the primacy" of Atatürk during his lifetime, in particular in terms of charisma: "During the years Atatürk was alive, Atatürk's authority, which derived from his charisma, was indispensable for the future of the regime in İnönü's eyes. It has been argued that Atatürk knew of İnönü's thinking on this matter and thus picked him as his Prime Minister in 1923." for example (Metin, 1998, p. 112).

45. Data primarily come from the following website: <http://www.ismetinonu.org.tr> and additional sources, described in Section C.1.

districts, with the highest administrative status and more likely to be targeted, as I do not know which districts İnönü targeted.

I find weak evidence that charisma mattered, on top of a visit, as shown in Table 3.8. Column (1) shows the baseline effect of a first visit, no matter who visited first, and confirms that the visits were efficient. They increased “Pure Turkish” first names by 0.53 percentage points in the full sample (Panel A) and by 0.74 percentage points in the restricted sample, excluding province centers (Panel B). Columns (2) and (3) display the effect when the district is visited by Atatürk for the first time, and Atatürk is the first to visit it or when the district is visited by İnönü for the first time, and İnönü is the first. Atatürk’s effect is larger than İnönü’s effect, which is not significant in the full sample. Column (4) shows the effect of the two treatments altogether: both leaders have an effect and İnönü also contributed to the increase in “Pure Turkish” names. Atatürk’s effect is larger than İnönü’s. The two coefficients, however are statistically indistinguishable: the p-value of the test of equality of coefficients equals 0.37 and 0.55 in the full and restricted samples respectively. Finally, Columns (5) and (6) display the differential effect of a visit by Atatürk (resp. İnönü) above a visit by any of them.

Overall, this finding suggests that leadership is an effective propaganda tool.

3.6.4 *Discussion: Indoctrination or Forced Assimilation?*

In this section, I discuss how to interpret the diffusion of “Pure Turkish” first names, in light of the results found. As already briefly mentioned in Section 3.3, there are three possible interpretations, given the repressive and authoritarian aspect of the regime (Kuran, 1995). First, giving a “Pure Turkish” first name could be a sign of sincere adherence to the reform program and to the new identity. Second, it could be a sign of believing that the regime is well-entrenched and will last: naming your child with a “Pure Turkish” name could therefore be opportunistic, to get access to new career opportunities and jobs for example. Finally, it could be the result of fear and forced assimilation, especially

among non-Turkish minorities urged to “Turkify” their culture.⁴⁶

At the same time, it is important to recall that there was no legal requirement to Turkify first names and the Kemalist state did not have the means to implement a heavy and totalitarian repression of its citizens—and therefore to systematically control and repress people not giving first names in “Pure Turkish” (Aymes et al., 2015; Yılmaz, 2013).⁴⁷ Many non-Turkish Muslims used very similar first names as Turks and could have simply kept using them.⁴⁸ Besides, the language reform happened in parallel with the Surname law, which legally asked citizens to pick new surnames—mostly among a set of “Pure Turkish” words. If individuals were fearing repression, they could adopt Pure Turkish last names—and indeed most of the population did (Türköz, 2018).

In an attempt to disentangle these interpretations, I run a difference-in-differences regression of the form given by equation 3.1, with the total number of minority villages from Nişanyan (2010) as heterogeneity variable. If forced assimilation was the main driver of the results, we would expect to see a high(er) take-up among minority localities. Results are displayed in Table A8, for the full sample (Panel A) and the restricted sample (Panel B). While there is no clear differences between places with a strong minority presence and others in Panel A, we observe a significant negative effect in places with a high number of Kurdish Villages (Column 4) and with a high number of minority villages (Column 5). This suggests that not only the main treatment is driven by Turkish areas, but the visits might have generated a slight backlash in minority places. Figure 3.8, displays the corresponding event-study figures, transforming the indicator for the strength of minority presence locally by a binary variable, indicating whether or not the

46. The language reform was indeed accompanied by a ban on the Kurdish language, the change in the names of minority towns and villages and systematic repression on ethno-religious minorities (Zeydanlıoğlu, 2012), which might have led to forced assimilation and fear among minorities.

47. The State’s view on how to treat of non-Muslim minorities with regard to their names is still subject to historiographical debates. There was probably a large variability in the concrete implementation of the reforms locally (Szurek, 2020).

48. As underlined in Aslan (2009): “Unlike non-Muslims, who conventionally used different names than the Muslims at the time, there was no distinct separation between Kurdish and Turkish names. Both ethnic communities used to give traditional Muslim names, which were predominantly Arabic and Persian, to their children.”, page 11.

district has a minority village. Results are qualitatively similar—although we do not observe a backlash effect anymore. The increase in “Pure Turkish” first names is mostly driven by Turkish areas. This finding suggests that adherence and opportunism played a role, and that the indoctrination was successful on average, among a specific segment of the population. This is consistent with my previous finding that the effect was driven by an educated elite. However, this does not rule out the fact that some people might have given “Pure Turkish” first name as a result of fear. As seen in Figure 3.8 (c), Kurdish areas actually take up more “Pure Turkish” names in the first years following a visit, but the effect does not last. Other minority areas take up in later years, once the regime is more established and that the propaganda institutionalized.

For all these reasons, the spread of new Pure Turkish first names in a given locality following a visit can be interpreted as an indicator of a more successful state control over the periphery, if not true adherence to Kemalism.

3.7 *Conclusion*

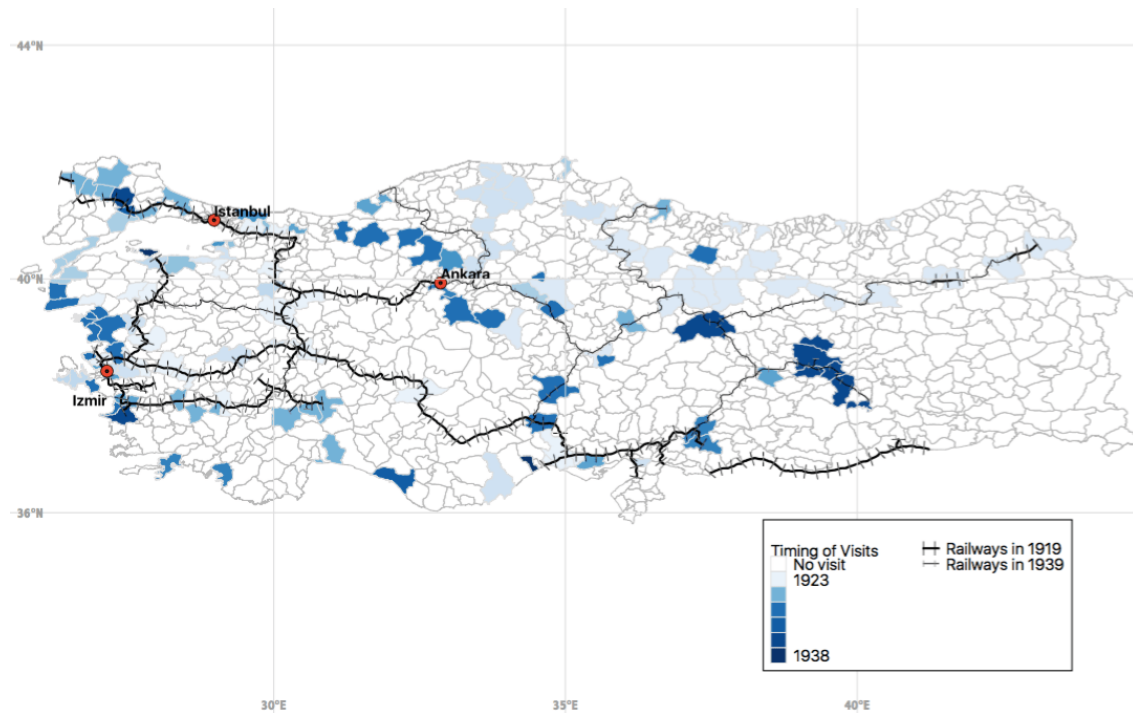
Can a charismatic leader contribute to the construction of a national identity? I answer this question by studying the role of Mustafa Kemal Atatürk in spreading the new Turkish language during the first years of the creation of Turkey. I assemble a novel and original historical database, with detailed information on his visits as well as other historical and geographic characteristics at the district level. I collect unique historical birth certificates and use first names in “Pure Turkish”, the new language introduced by the state as part of its nation-building effort, to measure the successful control of the central state over its periphery.

Using a difference-in-differences design that exploits time and geographic variation in Atatürk’s visits to districts, I show that visited districts are more likely to adopt first names in “Pure Turkish”. The effect is increasing over time, and is highest after fifteen years, when it then decreases and disappears. This result can mostly be explained by Atatürk’s ability to co-opt local elites. The visits also predict the opening of cultural branches of the ruling party, the “People’s Houses”, that in turn have a similar propaganda role and impact naming practices. This suggests that leader and institutions can be complement.

If my findings imply that a one-off exposure to a leader can an impact on identity, the effect, however, is relatively small, and the leader’s visits only explain 7 percent of the entire change, that is also due to other nation-building policies happening concurrently. First names, admittedly, provide only a partial window to understand historical change. Nevertheless, my results show that a leader can act as a coordination device and as a complement to future institutional changes, accelerating the change. This paper therefore constitutes, to my knowledge the first systematic evidence on the ability of an individual leader to contribute to the construction of a national identity. It also provides new empirical evidence to the old debate over the relative roles of individuals in shaping

historical outcomes and contributes to a large theoretical literature in economics that has extensively analyzed how leaders can make a difference, for better or worse.

Figure 3.1 – Number of Districts visited by Atatürk and Timing of the Visits



Notes: This figure depicts a map of Turkish contemporary districts (as of 2018), the main unit of observation of my analysis. Districts in blue are visited by Atatürk and districts in white are not. Districts in light blue were visited first, starting in 1923. Darker shades indicate districts visited later on, until 1938, the year of the last visit. Dark black lines indicate the railway network at the end of the Ottoman Empire, for the last year pre-treatment (1919). Lighter lines show the railway network after the visit period.

Sources: [Kocatürk \(1988\)](#) to identify visited districts and [Akgüngör et al. \(2011\)](#) for the railway networks

Figure 3.2 – Atatürk, the "Name Giver", "First Teacher" and "Father": Pictures taken during his Visits

(a)



(b)

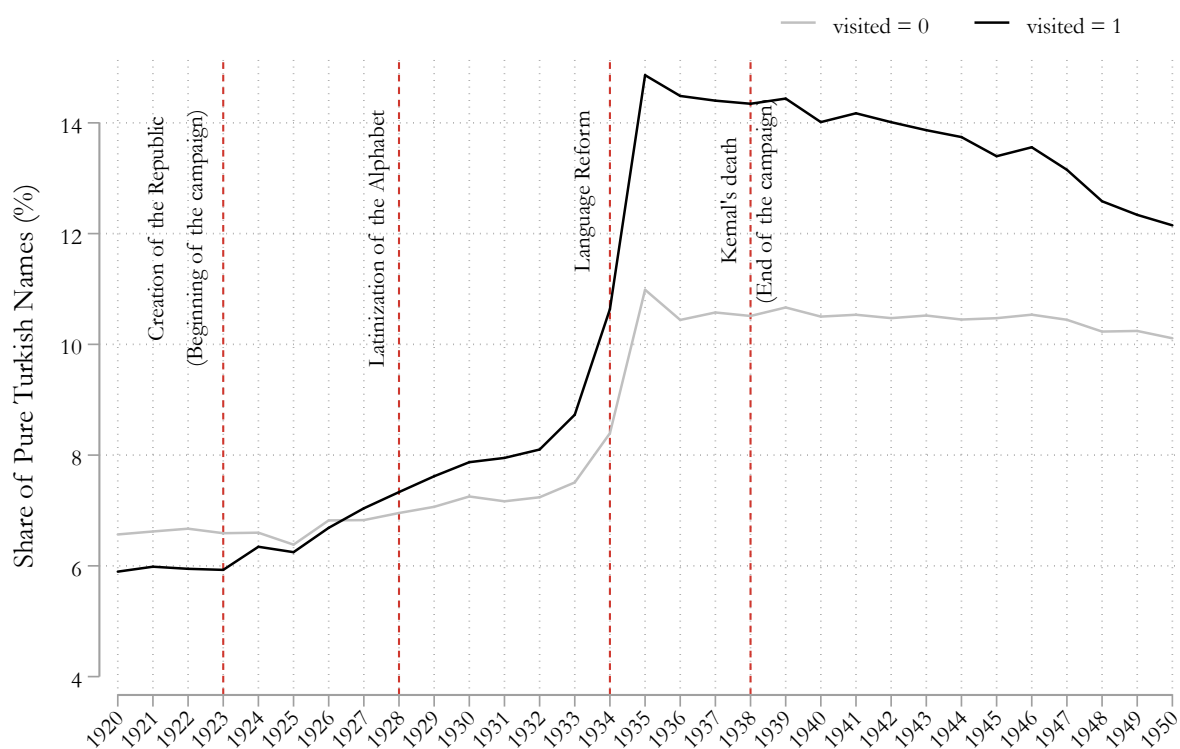


(c)



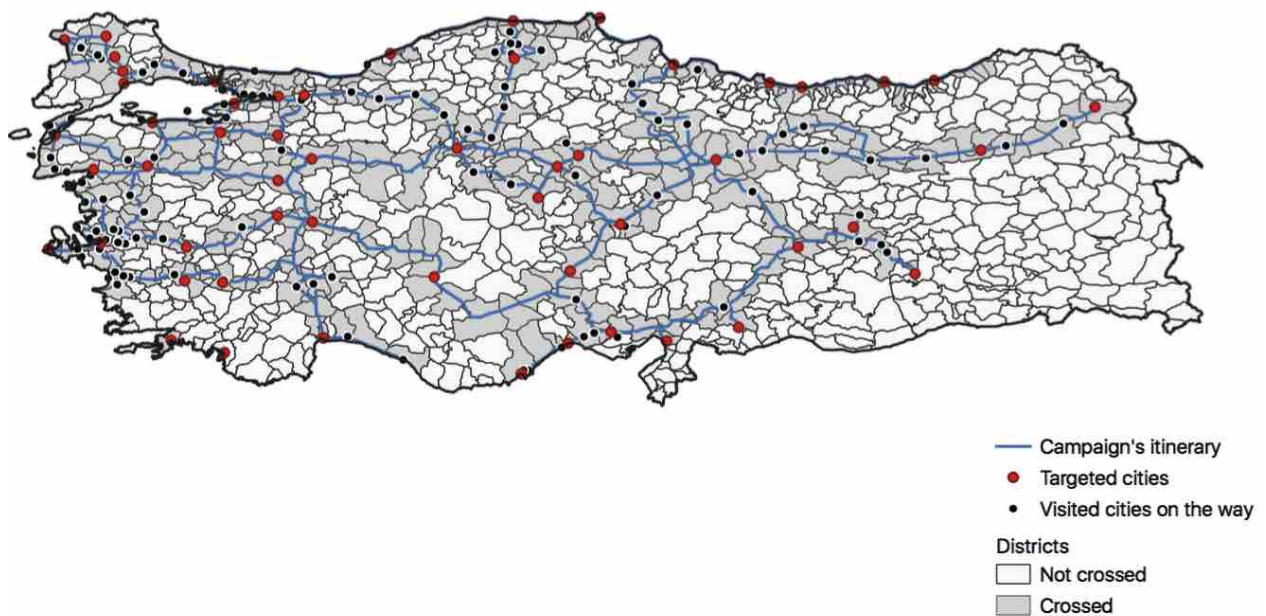
Source: Atatürk Research Center (Atatürk Araştırma Merkezi).

Figure 3.3 – Evolution of “Pure Turkish” First Names across Visited and Non-Visited districts



Notes: This figure shows the raw evolution of the share of Pure Turkish first names among newborns in visited and non-visited districts, for each year between 1920 and 1950. Sources: Population General Directorate for the historical birth certificates, [Türköz \(2018\)](#) for the list of “Pure Turkish” names; [Kocatürk \(1988\)](#) for the visit status.

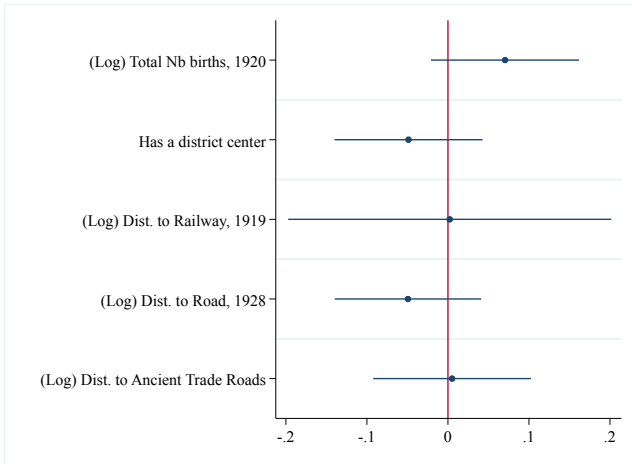
Figure 3.4 – Map of the Restricted Sample



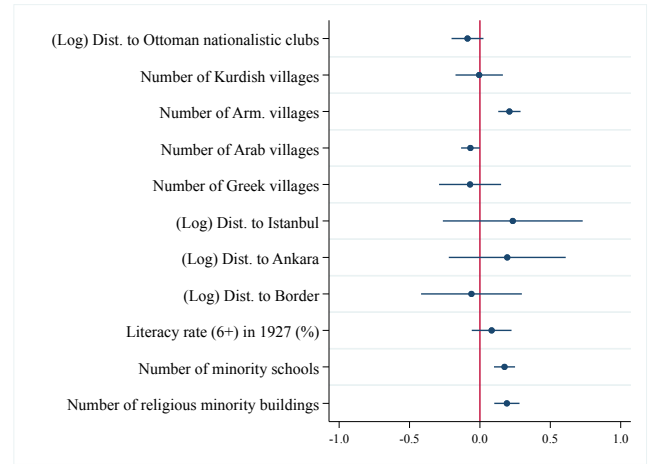
Notes: This figure displays the restricted sample used in my main empirical analysis. The red dots represent cities Atatürk targeted (the start and end points in his itinerary, or “termini” cities). The blue line represent his itinerary, computed using historical maps of the road and railway networks, information on his visits, topography data and the Least Cost Path feature in QGIS software. Grey districts are districts crossed at least once during the campaign effort. The restricted sample includes only the crossed districts, in grey, and excludes districts with a targeted “termini” city.

Figure 3.5 – Balance Plot between Visited and Non-Visited Districts (Restricted Sample)

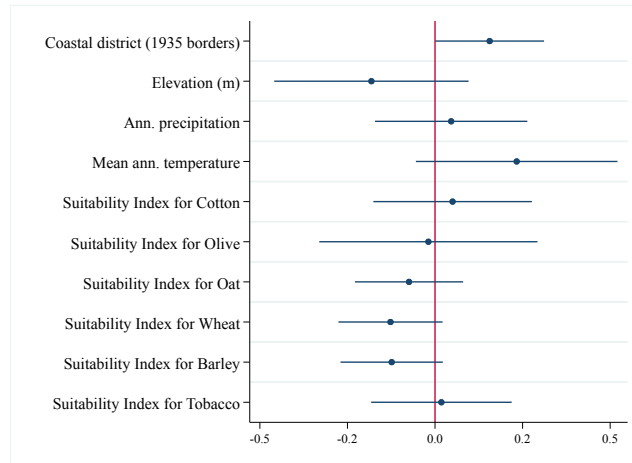
(a) Development and Demography



(b) Culture and Politics



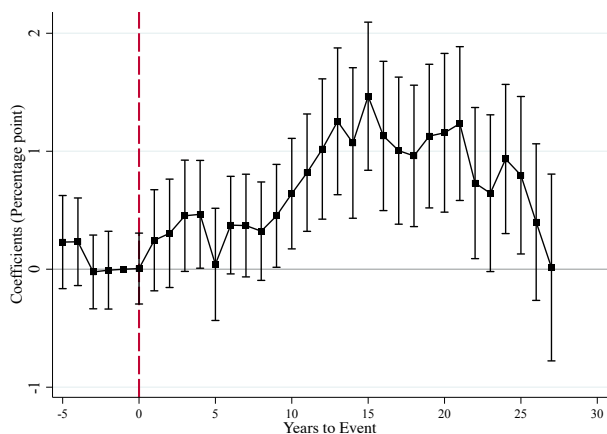
(c) Geography



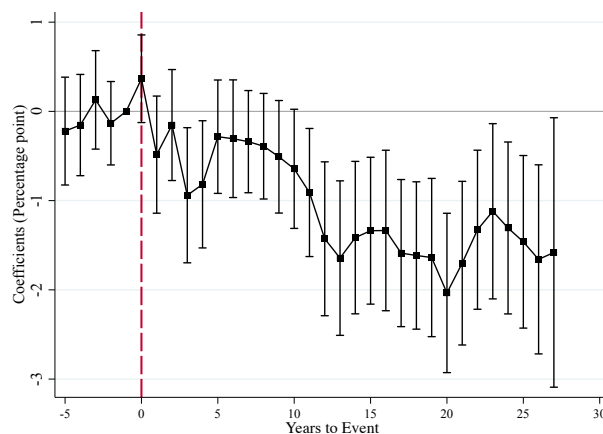
Notes: These graphs plot standardized beta coefficients of bi-variate regressions of the variables listed on the left on the visit status with province fixed effects and robust standard errors. Regressions are run using the restricted sample (removing targeted cities and focusing only on districts along his road) as displayed in Figure 3.4. Visited and non-visited districts along the road are comparable along various dimensions.

Figure 3.6 – Impact of Atatürk’s Visit on “Pure Turkish” and on Arabic First Names (Event-Study)

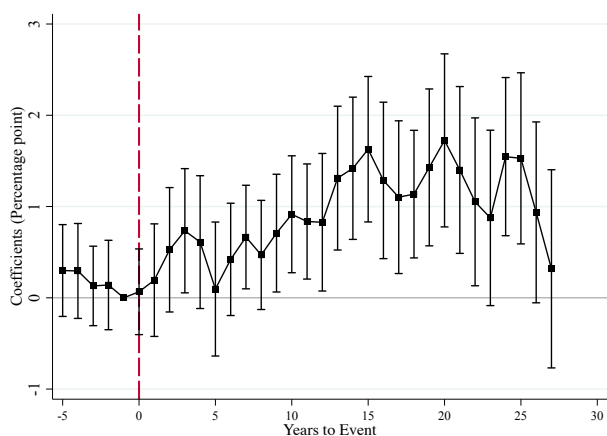
(a) Effect on Pure Turkish names, Full Sample



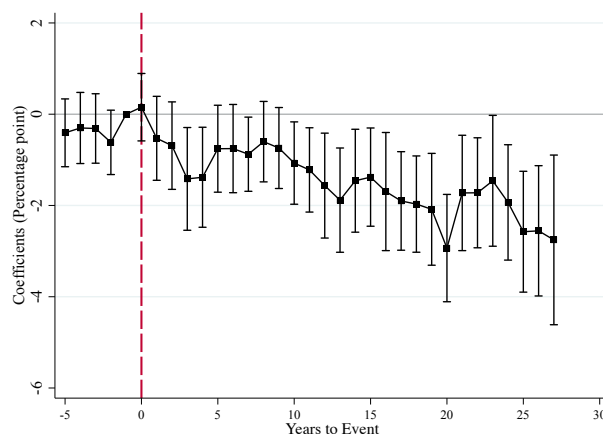
(b) Effect on Arab names, Full Sample



(c) Effect on Pure Turkish names, Restricted Sample

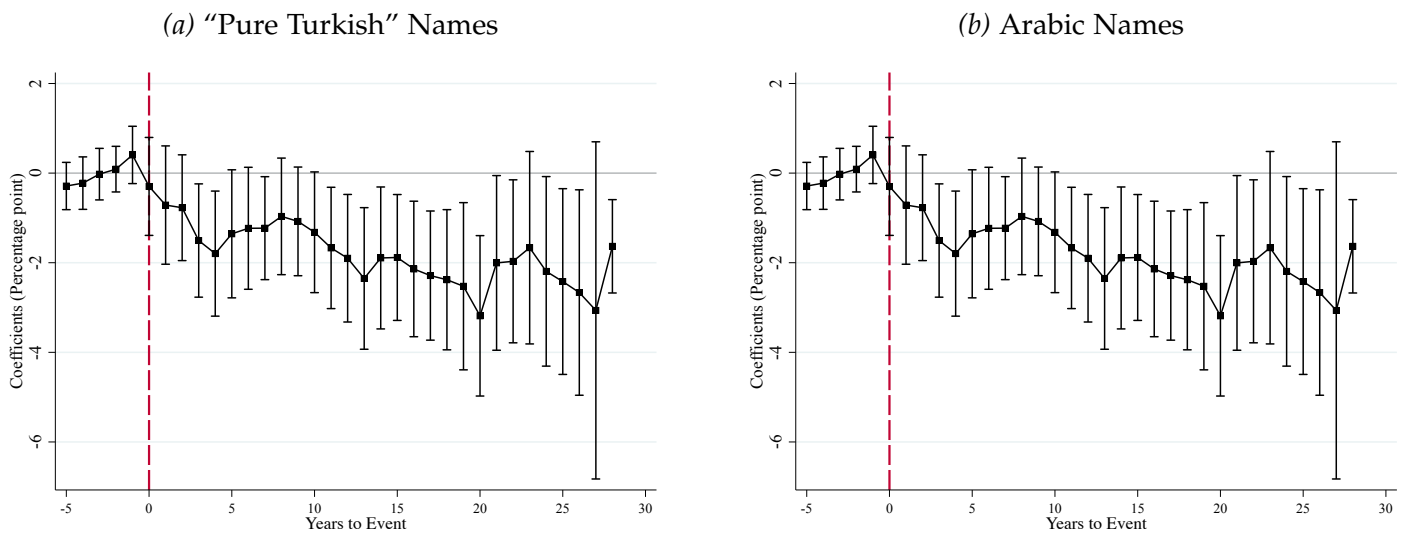


(d) Effect on Arab names, Restricted Sample



Notes: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2 run on both the full and the restricted sample, where the dependent variable is the share of newborns with Pure Turkish names or with Arabic names in a given district and year. The event is defined as the first time a district is visited by Atatürk. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

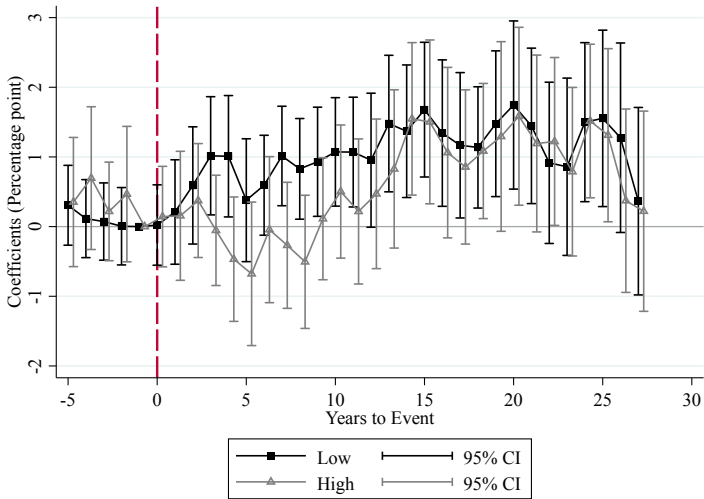
Figure 3.7 – De Chaisemartin- D’Haultfoeuille event-study results of the effect of Atatürk’s visits on names



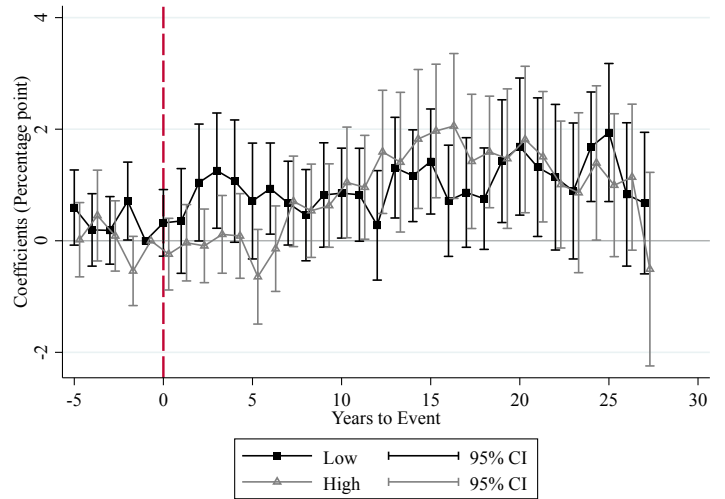
Notes: This figure presents event-study results of the effect of the visits on “Pure Turkish” and Arabic first names among newborns, using [de Chaisemartin and D’Haultfoeuille \(2020\)](#)’s method, implemented using the `did_multiplegt` command available on SSC repository.

Figure 3.8 – Heterogeneity Analysis, by Minority Presence

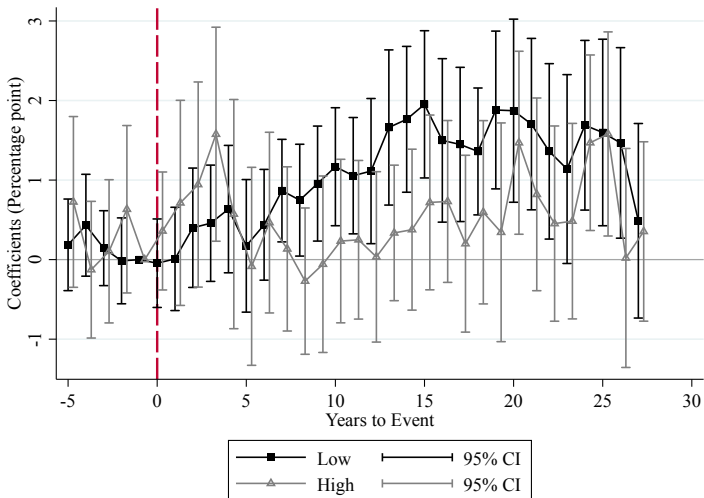
(a) Armenian Villages Density



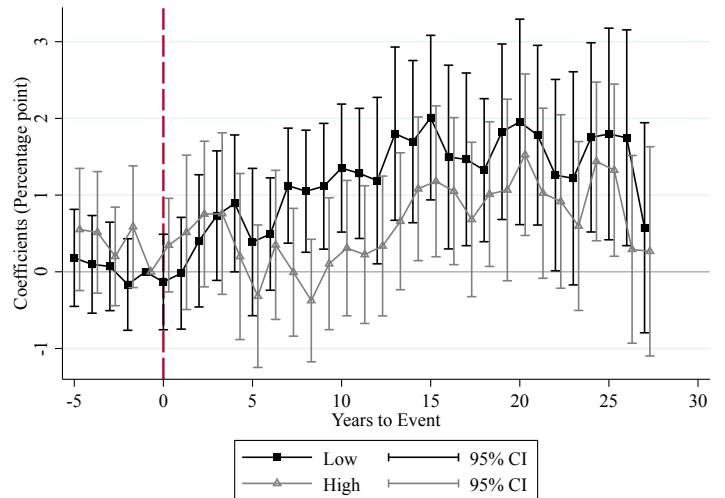
(b) Greek Villages Density



(c) Kurdish Villages Density

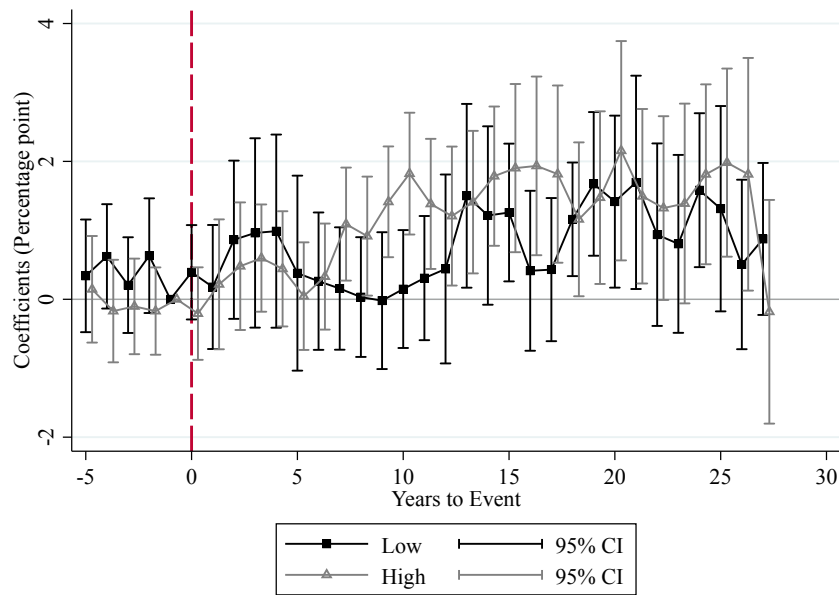


(d) Minority Villages Density



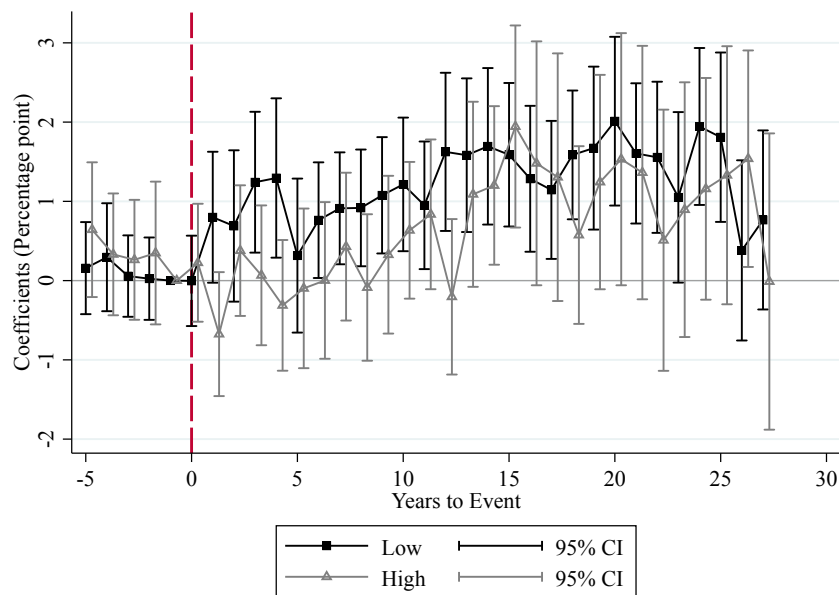
Notes: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2, with an interaction term between the right-hand-side visit variable and the heterogeneity variable of interest, which is a dummy variable indicating whether there is a high (or low) number of minority villages in a given district. Data on the villages come from Nisanyan (2010). The dependent variable is the share of Pure Turkish names in a given district and year. The event is defined as the first time a district is visited by Atatürk. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

Figure 3.9 – Heterogeneity Analysis, depending on whether or not Atatürk met with local elites



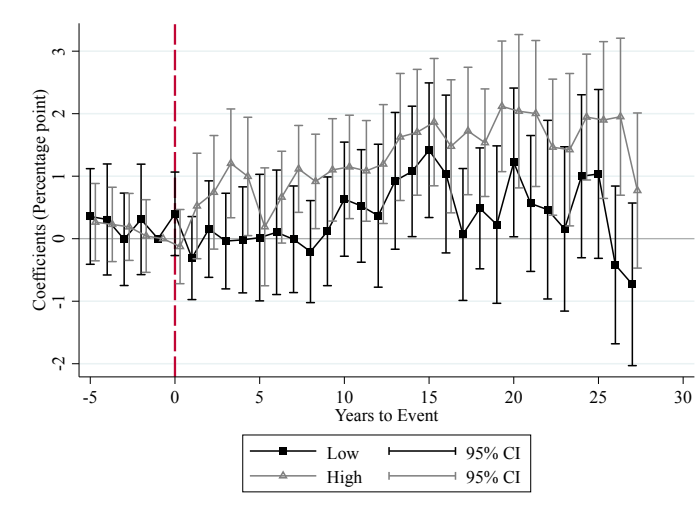
Notes: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2, with an interaction term between the right-hand-side visit variable and the heterogeneity variable of interest, which is a dummy variable indicating whether or not Atatürk met with local elites during the visit. The dependent variable is the share of Pure Turkish names in a given district and year. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

Figure 3.10 – Heterogeneity Analysis, by Distances to former Ottoman Nationalistic clubs



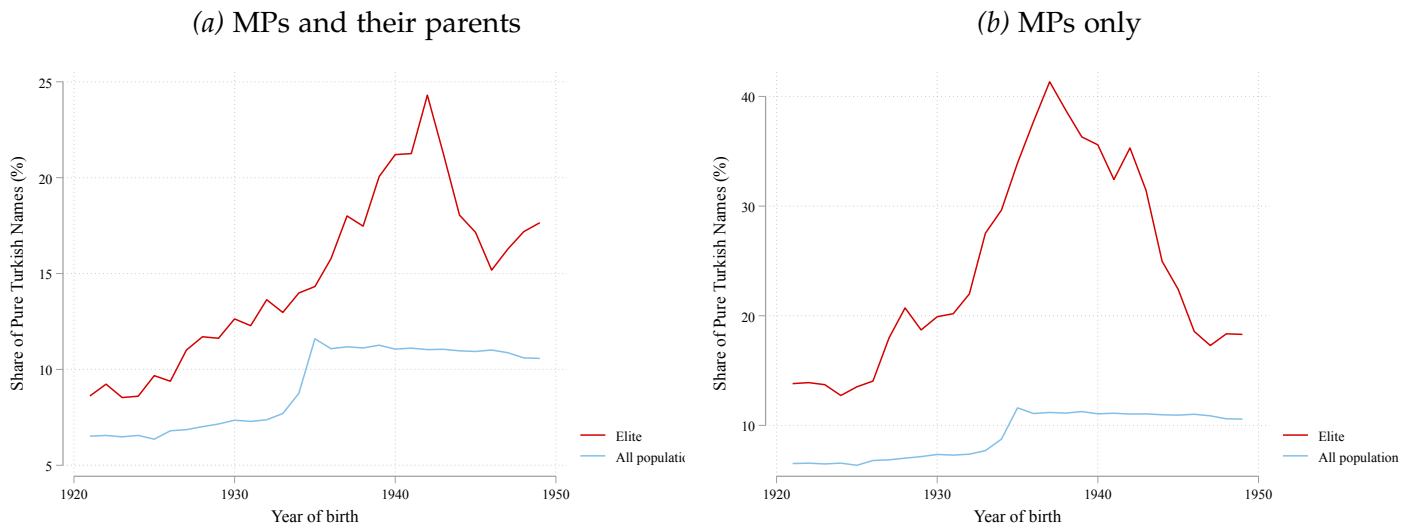
Notes: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2, with an interaction term between the right-hand-side visit variable and the heterogeneity variable of interest, which is a dummy variable indicating whether a district is below or above the median distance to the nearest Ottoman nationalistic club. The dependent variable is the share of Pure Turkish names in a given district and year. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

Figure 3.11 – Heterogeneity Analysis, by literacy rates



Notes: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2, with an interaction term between the right-hand-side visit variable and the heterogeneity variable of interest, which is a dummy variable indicating whether there is a high (or low) literacy rates in 1927 (above or below the median). Data on the literacy rates come from the 1927 census. The dependent variable is the share of Pure Turkish names in a given district and year. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

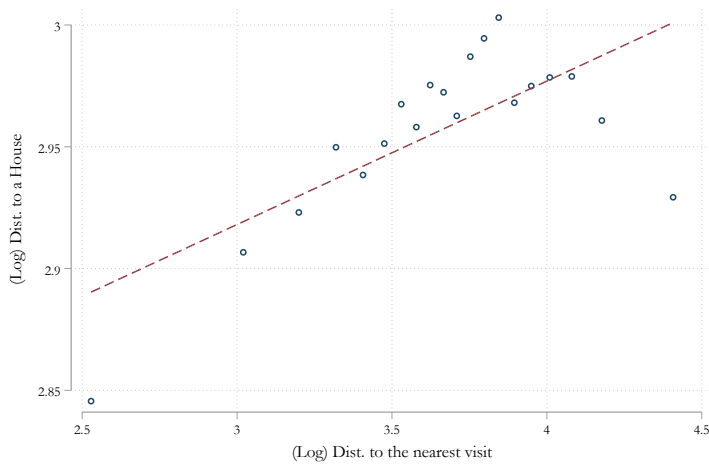
Figure 3.12 – Evolution of the share of Pure Turkish names among the elite and among all newborns in Turkey, 1920-1950



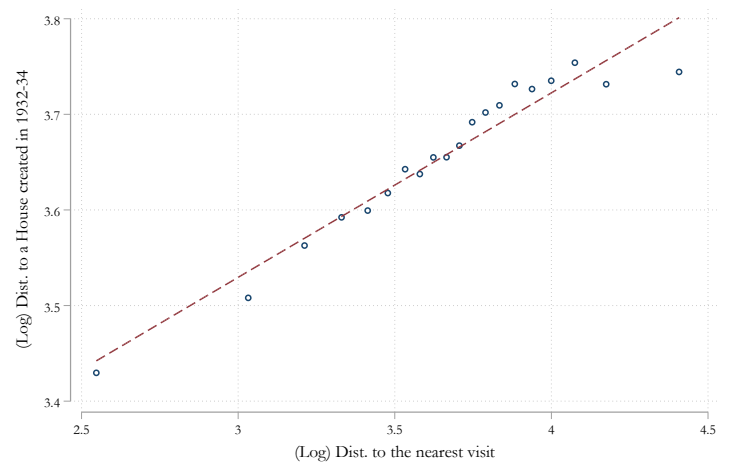
Notes: This figure plots the evolution of the share of Pure Turkish names among deputy members and their parents using the bibliographical data from the Library of the Grand National Assembly and compares it to the evolution of the share of Pure Turkish names among all newborns using the universe of birth certificates, between 1920 and 1950.

Figure 3.13 – Correlation between the Distance to the nearest visited city and the nearest House

(a) All People's Houses



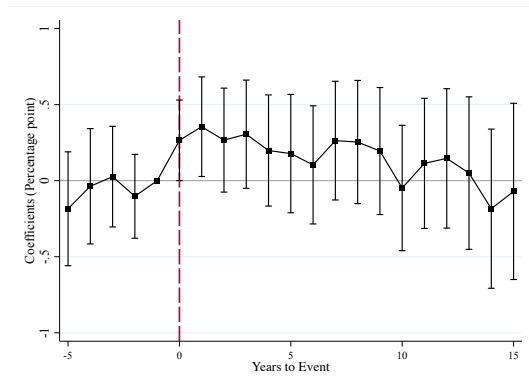
(b) Houses opened between 1932-1934



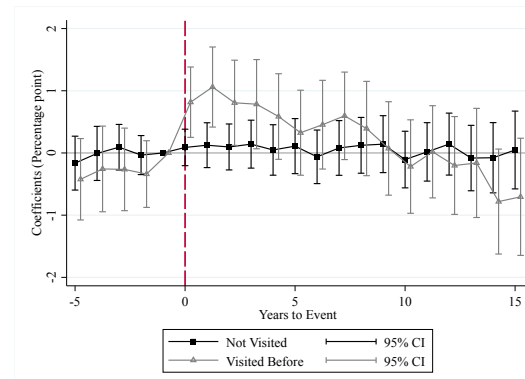
Notes: These figures display binscatter plots of the distance to the nearest visits and the distance to the nearest House, with district fixed effects and covariates selected using a Lasso procedure. The visits and the houses are highly correlated.

Figure 3.14 – Impact of the Opening of a People’s House on Pure Turkish and Kurdish First Names and Heterogeneity depending on the Visit Status

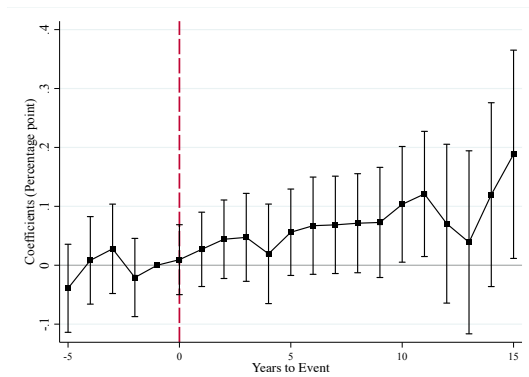
(a) Effect of the Houses on Pure Turkish Names



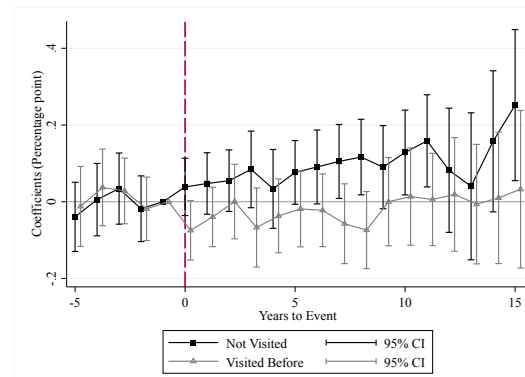
(b) Effect of the Houses on Pure Turkish Names- Heterogeneity by visit status



(c) Effect of the Houses on Kurdish Names



(d) Effect of the Houses on Kurdish Names- Heterogeneity by visit status



Notes: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2 run on the restricted sample, where the dependent variable is the share of Pure Turkish names (a) or of Kurdish names (b) among newborns, in a given district and year. The event is defined as the first time a People’s House is opened in a given district and year. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

Table 3.1 – Ten most frequent first names in “Pure Turkish” and in Arabic or Turkish among newborns in 1920 and in 1940

1920					1940				
Name	Ranking	# Individ.	Freq.	Cum. Freq.	Name	Ranking	# Individ.	Freq.	Cum. Freq.
PANEL A: PURE TURKISH NAMES									
Gulsum	33	947	.35	.35	Yasar	20	3163	.53	.55
Kazim	43	794	.29	.65	Sevim	23	2879	.485	1
Dursun	54	655	.24	.9	Dursun	34	1983	.335	1.35
Hakki	72	559	.205	1.1	Bayram	50	1536	.255	1.6
Durmus	79	516	.19	1.3	Yilmaz	53	1518	.255	1.85
Yasar	84	498	.185	1.45	Kazim	54	1515	.255	2.1
Bayram	91	444	.165	1.6	Gulsum	56	1498	.25	2.35
Sati	94	433	.16	1.8	Yuksel	59	1418	.24	2.6
Elmas	99	421	.155	1.95	Turkan	68	1294	.215	2.8
Sefer	107	403	.15	2.1	Ayten	69	1290	.215	3.05
PANEL B: ARABIC OR TURKISH NAMES									
Mehmet	1	14735	5.415	5.4	Mehmet	1	24072	4.035	4.05
Fatma	2	13615	5.005	10.4	Fatma	2	19222	3.225	7.25
Ayse	3	9261	3.405	13.8	Ali	3	14247	2.39	9.65
Ali	4	7800	2.865	16.7	Ayse	4	14212	2.385	12.05
Ahmet	5	7758	2.85	19.55	Mustafa	5	14195	2.38	14.4
Mustafa	6	7745	2.845	22.4	Ahmet	6	12509	2.1	16.5
Emine	7	7145	2.625	25	Emine	7	11395	1.91	18.4
Hasan	8	6128	2.25	27.25	Hasan	8	10558	1.77	20.2
Hatice	9	5788	2.125	29.4	Huseyin	9	10192	1.71	21.9
Huseyin	10	5777	2.125	31.5	Hatice	10	9010	1.51	23.4

Notes: The ten most frequent first names “Pure Turkish” and in Arabic or Turkish, given in 1920 and in 1940 in the birth certificates. The frequency and cumulative frequency (in percentage) are computed relative to the entire population of newborns: 0.35 percent of the babies born in 1920 were given the name Gulsum and 5.4 percent were given the name Mehmet. Overall, the ten most frequently given “Pure Turkish” names account for 2.1 percent of the total population fo newborns in 1920, while the ten most frequently given Arabic and Turkish names account for 30 percent.

Table 3.2 – Predicting the visits: Probit estimates of Ataturk’s appearances, by year

Dep. Var:	Visited between 1923-1938		Visited between 1923-1924		Visited between 1925-1928		Visited between 1929-1933		Visited between 1934-1938	
	Coeff.	(S.E.)	Coeff.	(S.E.)	Coeff.	(S.E.)	Coeff.	(S.E.)	Coeff.	(S.E.)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Demography and Development										
(Log) Total Nb births, 1920	0.530***	(0.113)	0.636***	(0.198)	0.577**	(0.248)	0.410***	(0.132)	0.245*	(0.145)
Has a province center	1.938***	(0.255)	1.363***	(0.253)	0.215	(0.385)	0.855***	(0.295)	0.120	(0.347)
Has a district center	0.233	(0.164)	0.235	(0.230)	0.476	(0.342)	-0.0446	(0.250)	0.0640	(0.208)
(Log) Dist. to Railway, 1919	-0.0631	(0.0619)	-0.264***	(0.0807)	0.0230	(0.111)	0.0348	(0.0752)	0.0579	(0.0902)
(Log) Dist. to Road, 1928	-0.120**	(0.0553)	-0.112	(0.0715)	0.0590	(0.104)	0.0161	(0.0769)	-0.170**	(0.0714)
(Log) Dist. to Ancient Trade Roads	-0.0242	(0.0528)	-0.0667	(0.0664)	0.151	(0.114)	-0.0576	(0.0749)	-0.00461	(0.0709)
City density	84.82**	(39.05)	60.16*	(32.41)	44.66	(32.72)	58.36**	(25.84)	85.98***	(28.23)
Culture and Politics										
(Log) Dist. to Ottoman nationalistic clubs	-0.182**	(0.0886)	0.152	(0.107)	-0.327**	(0.132)	-0.183	(0.115)	0.0228	(0.122)
Density of Kurdish villages	0.0812	(0.123)	-0.639**	(0.273)	0.437	(0.375)	0.0602	(0.325)	0.0724	(0.126)
Density of Arm. villages	0.210	(0.236)	0.268	(0.215)	-7.091**	(3.366)	-1.303	(1.111)	0.414	(0.282)
Density of Arab villages	-13.28***	(4.506)	-10.04	(8.580)			-4.409	(4.454)	-4.023**	(1.980)
Density of Greek villages	-0.0143	(0.0728)	-0.0315	(0.0763)	0.0633	(0.108)	-0.0669	(0.0665)	0.143	(0.140)
(Log) Dist. to Istanbul	0.271**	(0.118)	0.274	(0.180)	0.417	(0.329)	0.00141	(0.150)	-0.0789	(0.198)
(Log) Dist. to Ankara	-0.262*	(0.148)	0.518*	(0.281)	-0.491**	(0.198)	-0.479*	(0.259)	-0.182	(0.189)
(Log) Dist. to Border	0.229**	(0.0955)	0.655***	(0.193)	0.0213	(0.200)	0.101	(0.128)	0.123	(0.106)
Occupied after WWI	0.0436	(0.226)	0.511*	(0.308)	0.315	(0.471)	-0.266	(0.276)	-0.436	(0.350)
Density of minority schools	0.487	(2.744)	6.712*	(3.942)	-18.34	(11.73)	21.09*	(12.06)	1.894**	(0.792)
Density of religious minority buildings	-0.00565	(0.0110)	-0.0221*	(0.0118)	0.0230*	(0.0135)	-0.133**	(0.0603)	-0.0188***	(0.00521)
Geography										
Coastal	0.972***	(0.233)	0.526	(0.348)	1.270***	(0.409)	0.264	(0.291)	0.659**	(0.320)
Ann. precipitation (cm)	-0.904	(0.575)	0.0212	(0.734)	-2.516*	(1.285)	0.985*	(0.537)	-1.494	(1.130)
Mean ann. temperature	-0.0994	(0.0633)	-0.191	(0.138)	-0.251*	(0.138)	-0.123	(0.127)	-0.0292	(0.0819)
Elevation (km)	-0.880***	(0.320)	0.290	(0.445)	-1.155*	(0.679)	-0.442	(0.395)	-1.225***	(0.470)
Suitability Index for Cotton	2.509	(3.386)	2.877	(6.287)	-3.843	(6.111)	7.517	(4.825)	2.751	(4.593)
Suitability Index for Olive	0.947*	(0.526)	0.121	(0.779)	1.906*	(1.021)	1.793***	(0.645)	0.0852	(0.732)
Suitability Index for Oat	0.952	(1.354)	1.992	(2.096)	7.431**	(2.998)	-2.088	(1.743)	-0.637	(1.614)
Suitability Index for Wheat	-1.741	(1.065)	0.411	(1.313)	-7.415	(6.921)	0.562	(1.501)	-48.81*	(25.15)
Suitability Index for Barley	1.668**	(0.831)	-0.461	(0.687)	6.430	(6.799)	-0.538	(1.020)	48.14*	(24.79)
Suitability Index for Tobacco	-2.227***	(0.710)	1.061	(1.230)	-4.090***	(1.498)	1.139	(1.626)	-1.881**	(0.931)
Suitability Index for Potato	0.00286	(0.0106)	0.0152	(0.0166)	0.00760	(0.0206)	-0.00960	(0.0134)	-0.00229	(0.0164)
Constant	-2.772	(1.812)	-14.82***	(3.456)	-6.201	(4.214)	1.317	(2.554)	1.875	(2.148)
Observations/ # Visited cities	973/153		973/52		973/28		973/36		973/37	

Notes: *** p<0.01, ** p<0.05, * p<0.1. This table reports probit estimates and their standard errors, clustered at the district level. Column (1) reports the estimates for all visits (between 1923 and 1938); Column (3), (5), (7) and (9) report the estimates for the visits for different time periods.

Table 3.3 – Main Results: Effect of Atatürk’s Visits on Naming Practices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DEPENDENT VARIABLES: SHARE OF NEWBORNS WITH A:							
	NEW NAMES		TRADITIONAL NAMES		MINORITY NAMES		
	Pure Turkish Name	Arabic Name	Religious Name	Muhammed	Kurdish Name	Armenian Name	Jewish Name
Panel A: Full Sample							
Visited × Post	0.436** (0.203)	-0.616** (0.286)	-0.056 (0.065)	-0.024** (0.012)	0.003 (0.040)	-0.009 (0.007)	-0.005 (0.004)
Observations	29760	29760	29760	29760	29760	29760	29760
R-squared	0.750	0.813	0.671	0.444	0.934	0.644	0.783
Mean of outcome at baseline	6.523	71.44	3.386	0.0563	1.467	0.0624	0.0246
s.d. of outcome	3.432	9.142	2.866	0.269	3.178	0.219	0.124
Panel B: Restricted Sample							
Visited × Post	0.556** (0.263)	-0.776** (0.385)	-0.014 (0.083)	-0.016 (0.013)	-0.023 (0.050)	-0.006 (0.011)	-0.003 (0.004)
Observations	8432	8432	8432	8432	8432	8432	8432
R-squared	0.796	0.771	0.482	0.308	0.833	0.589	0.727
Mean of outcome at baseline	5.545	74.38	2.839	0.0383	0.476	0.0498	0.0211
s.d. of outcome	0.785	6.273	1.953	0.204	1.098	0.181	0.103
Year FE	✓	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The Table presents the results of the estimation of specification 3.1 on the full sample (Panel A) and on the restricted Sample (Panel B). The dependent variables are the share of first names by type (Pure Turkish, Arabic, Religious or minority first names) among newborns. The unit of observation is the district. Baseline controls are described in section 3.4.3. Standard errors in parenthesis, clustered at the district level, the level of the treatment. A visit increases the share of Pure Turkish names, and decreases the share of Arabic names. It has no effect on other religious names and on (non-Turkish) minority names (Kurdish, Armenian and Jewish names).

Table 3.4 – Effect of Atatürk’s Visits on Naming Practices—Additional effect of a year following a visit

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DEPENDENT VARIABLES: SHARE OF NEWBORNS WITH A:							
	NEW NAMES		TRADITIONAL NAMES		MINORITY NAMES		
	Pure Turkish Name	Arabic Name	Religious Name	Muhammed	Kurdish Name	Armenian Name	Jewish Name
Panel A: Full Sample							
Visited × Years since the Visit	0.036*** (0.013)	-0.059*** (0.018)	-0.008* (0.005)	-0.003*** (0.001)	0.001 (0.003)	-0.001 (0.000)	-0.000 (0.000)
Observations	29760	29760	29760	29760	29760	29760	29760
R-squared	0.750	0.813	0.671	0.444	0.934	0.644	0.783
Mean of outcome at baseline	6.523	71.44	3.386	0.0563	1.467	0.0624	0.0246
s.d. of outcome	3.432	9.142	2.866	0.269	3.178	0.219	0.124
Panel B: Restricted Sample							
Visited × Years since the Visit	0.053*** (0.017)	-0.078*** (0.024)	-0.006 (0.006)	-0.002* (0.001)	0.002 (0.003)	-0.000 (0.001)	-0.000 (0.000)
Observations	8432	8432	8432	8432	8432	8432	8432
R-squared	0.796	0.772	0.482	0.308	0.833	0.589	0.727
Mean of outcome at baseline	5.545	74.38	2.839	0.0383	0.476	0.0498	0.0211
s.d. of outcome	0.785	6.273	1.953	0.204	1.098	0.181	0.103
Year FE	✓	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The Table presents the results of the estimation of specification 3.1 but using as treatment variable a variable which equals zero if the district is not visited or not yet visited, and which equals the years since the visits when visited. Panel A displays the results using the full sample and Panel B using the restricted sample. The dependent variables are the share of first names by type (Pure Turkish, Arabic, Religious or minority first names) among newborns. The unit of observation is the district. Baseline controls are described in section 3.4.3. Standard errors in parenthesis, clustered at the district level, the level of the treatment. The average increase in “Pure Turkish” names in visited districts in the restricted sample equals 0.05 percentage points for any year following a visit.

Table 3.5 – The Visits are among the main predictors of the Opening of a People’s House

Dep. Variables: Model Predictors	(1)		(2)		(3)	
	(Log) Dist. House		(Log) Dist. House		Has a House	
	OLS		OLS		Probit	
	All		Selected by LASSO		Selected by LASSO	
	Coeff.	S.E.	Coeff.	S.E	Coeff.	S.E
(Log) Dist. to the nearest Visit	0.08***	(0.01)	0.08***	(0.01)	-0.64***	(0.14)
(Log) Dist. to the nearest Prov center	0.33***	(0.02)	0.34***	(0.02)	-0.86**	(0.41)
(Log) Dist. to the nearest Dis. center	0.63***	(0.01)	0.63***	(0.01)	-3.27***	(0.10)
(Log) Dist. to the road, 1928	0.05***	(0.00)	0.04***	(0.00)	-0.26***	(0.06)
(Log) Dist. to the rail, 1919	0.01	(0.01)				
(Log) Dist. to the rail, 1932	-0.02***	(0.01)				
(Log) Dist. to Ott. nationalistic club	0.04***	(0.01)	0.04***	(0.01)	0.91***	(0.18)
(Log) Dist. to the border	0.08***	(0.03)				
(Log) Dist. to Ankara	-0.53***	(0.05)	-0.52***	(0.05)	0.79	(1.69)
(Log) Dist. to Istanbul	0.22***	(0.06)				
(Log) Dist. to the nearest rebellion	0.03***	(0.01)	0.03***	(0.01)	-0.69***	(0.23)
(Log) Dist. to the nearest battle	0.09***	(0.01)	0.10***	(0.01)	-0.92***	(0.16)
Nb of arab villages within 10 km	0.02***	(0.00)	0.02***	(0.00)	0.19***	(0.07)
Nb of arm villages within 10 km	-0.00	(0.00)				
Nb of greek villages within 10 km	-0.00***	(0.00)				
Nb of kurd villages within 10 km	-0.00	(0.00)				
Nb of arab villages within 20 km	0.01***	(0.00)	0.01***	(0.00)	-0.00	(0.04)
Nb of arm villages within 20 km	-0.00	(0.00)				
Nb of greek villages within 20 km	0.00	(0.00)				
Nb of kurd villages within 20 km	-0.00***	(0.00)	-0.00***	(0.00)	-0.03*	(0.02)
(Log) Dist. to the nearest Arab village	-0.21***	(0.03)	-0.23***	(0.03)	-0.57	(1.05)
(Log) Dist. to the nearest Arm. village	-0.35***	(0.04)	-0.28***	(0.04)	-5.43***	(1.92)
(Log) Dist. to the nearest Greek village	-0.47***	(0.07)	-0.27***	(0.04)	2.71**	(1.08)
(Log) Dist. to the nearest Kurdish village	0.00	(0.00)				
(Log) Dist. to the shore	0.06*	(0.03)				
Annual precipitation	0.01***	(0.00)				
Frost free period	0.00***	(0.00)				
Growing Period Length	-0.00**	(0.00)	-0.00***	(0.00)	0.01**	(0.00)
Annual Temperature	-0.00	(0.00)				
Elevation	0.00***	(0.00)	0.00***	(0.00)	-0.00***	(0.00)
Growing period length	0.00***	(0.00)				
Growing period mean temperature	-0.00***	(0.00)				
Suitability index for cotton	0.00***	(0.00)				
Suitability index for oat	-0.00**	(0.00)	-0.00***	(0.00)	0.00	(0.00)
Suitability index for olive	-0.00***	(0.00)	-0.00**	(0.00)	-0.01***	(0.00)
Suitability index for tobacco	0.03**	(0.01)				
Suitability index for wheat	0.00	(0.00)				
Suitability index for barley	-0.00	(0.00)				
Constant	2.29***	(0.26)	3.05***	(0.23)	19.97**	(8.42)
Observations	35,614		35,703		17,487	
R-squared	0.841		0.841			
District FE	✓		✓		✓	

Notes: *** p<0.01, ** p<0.05, * p<0.1. This table shows the effect of the logarithm of the distance to the nearest visited on the distance to the nearest house (Columns 1 and 2) and on the probability on having a house (Column 3). Columns 1 and 3 include a large set of historical and geographic covariates; Column 2 include predictors selected using a Lasso procedure. District fixed effects included and robust standard errors in parentheses.

Table 3.6 – Effect of the People’s Houses on first names and Heterogeneity depending on the visit status of the district

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Dep. Var:</i>	SHARE OF NEWBORNS WITH A:											
	Pure Turkish Name			Arabic Name			Kurdish Name			Religious Name		
Has a House × Post	0.538*** (0.168)	0.536*** (0.168)	0.435** (0.179)	-0.599** (0.234)	-0.595** (0.233)	-0.483** (0.245)	0.079* (0.044)	0.079* (0.044)	0.101** (0.047)	-0.028 (0.055)	-0.028 (0.054)	-0.016 (0.058)
Visited × Post Visit		0.357* (0.206)			-0.528* (0.292)			-0.003 (0.042)			-0.043 (0.067)	
Has a House × Post × Visited Before			0.574** (0.280)			-0.644 (0.415)		-0.122* (0.066)	-0.067 (0.066)			-0.035** (0.080)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760
Mean of outcome	6.523	6.523	6.523	71.44	71.44	71.44	1.467	1.467	1.467	3.386	3.386	3.386
s.d. of outcome	3.432	3.432	3.432	9.142	9.142	9.142	3.178	3.178	3.178	2.866	2.866	2.866

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table presents the results of the estimation of specification similar to equation 3.1, but where the treatment variable is a dummy variable switching to one the first year a People’s House is established in a given district. The dependent variables are the share of first names by type (Pure Turkish, Arabic, Religious or minority first names). The unit of observation is the district. Baseline controls included. Standard errors in parenthesis, clustered at the district level, the level of the treatment.

Table 3.7 – Effect of Atatürk’s visits, by type of activity

<i>Dep. Var: Sh. of newborns with a</i>	(1)	(2)	(3)	(4)
	"PURE TURKISH" NAME			
Panel A: Full Sample				
Visited × Post	0.45** (0.20)	0.20 (0.36)	0.40 (0.36)	0.25 (0.24)
Visited × Post × Elite		0.26 (0.42)		
Visited × Post × Mass			-0.07 (0.45)	
Visited × Post × Speech				0.81 (0.52)
Observations	29760	29078	29078	29078
R-squared	0.751	0.748	0.748	0.748
Mean of outcome	9.086	9.058	9.058	9.058
s.d. of outcome	4.456	4.438	4.438	4.438
Panel B: Restricted Sample				
Visited × Post	0.54** (0.26)	0.46 (0.46)	1.28*** (0.45)	0.88*** (0.33)
Visited × Post × Elite		0.90* (0.54)		
Visited × Post × Mass			-0.68 (0.61)	
Visited × Post × Speech				0.56 (0.88)
Observations	8432	8060	8060	8060
R-squared	0.795	0.779	0.779	0.779
Mean of outcome	9.324	9.319	9.319	9.319
s.d. of outcome	4.348	4.371	4.371	4.371
Year FE	✓	✓	✓	✓
District FE	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table shows results from regressions similar to the one shown in equation 3.1, where the dependent variable is the share of Pure Turkish names but where the treatment variables vary. Visited × Post is the same treatment variable as in equation 3.1. Elite is a dummy variable indicating whether Atatürk met with local elites in a given district. Mass is a dummy variable indicating whether he met with the crowd. Speech is a dummy variable indicating whether Atatürk made a speech. All regressions are run on the main restricted sample. Standard errors are clustered at the district level.

Table 3.8 – Effect of Atatürk’s vs İnönü’s Visits on First Names in Pure Turkish

<i>Dep. Var: Share of newborns with a</i>	(1)	(2)	(3)	(4)	(5)	(6)
	"PURE TURKISH" NAME					
Panel A: Full Sample						
Visited × Post (no matter by whom)	0.53*** (0.15)				0.48** (0.19)	0.74*** (0.24)
Visited × Post × Atatürk, and Atatürk is first		0.52** (0.22)		0.56** (0.22)	0.11 (0.28)	
Visited × Post × İnönü, and İnönü is first			0.26 (0.18)	0.32* (0.18)		-0.39 (0.28)
<i>P-value of the test of equality of coefficients</i>				0.37		
Observations	29760	29760	29760	29760	29760	29760
R-squared	0.750	0.750	0.750	0.750	0.750	0.750
Nb. Treated districts	160	96	55	152	96	55
Panel B: Excluding Province Centers						
Visited × Post (no matter by whom)	0.58*** (0.16)				0.56*** (0.20)	0.75*** (0.27)
Visited × Post × Ataturk, and Ataturk is first		0.54** (0.25)		0.58** (0.25)	0.04 (0.31)	
Visited × Post × Inonu, and Inonu is first			0.36** (0.17)	0.40** (0.17)		-0.33 (0.31)
<i>P-value of the test of equality of coefficients</i>				0.55		
Observations	27993	27993	27993	27993	27993	27993
R-squared	0.736	0.735	0.735	0.736	0.736	0.736
Nb. Treated	116	68	45	142	68	45
Year FE	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table shows the estimation results of difference-in-differences model of the form given by equation 3.1, where the dependent variable is the share of Pure Turkish names but where the treatment variables vary. In Column (1), the treatment variable is a dummy switching to one the first time a district is visited, either by Atatürk or by İnönü, and which stays equal to one. In Column (2), the treatment variable is a dummy equal to one the first time a district is visited, and Atatürk is the first to visit it. In Column (3), the treatment variable is a dummy equal to one the first time a district is visited by İnönü, and İnönü is the first to visit it. Panel A focuses on the full sample, while Panel B exclude province centers, the most populated districts with the highest administrative status. Standard errors are clustered at the district level. Robust standard errors in parentheses.

Bibliography

- Abramitzky, R., Boustan, L. P., and Eriksson, K. (2016). Cultural Assimilation during the Age of Mass Migration. NBER Working Papers 22381, National Bureau of Economic Research, Inc.
- Acemoglu, D., Fergusson, L., Robinson, J. A., Romero, D., and Vargas, J. F. (2016). The perils of high-powered incentives: Evidence from Colombia's false positives. Working Paper 22617, National Bureau of Economic Research.
- Acemoglu, D. and Jackson, M. (2015). History, Expectations, and Leadership in the Evolution of Social Norms. *Review of Economic Studies*, 82(2):423–456.
- Acemoglu, D. and Robinson, J. A. (2021). Culture, Institutions and Social Equilibria: A Framework. NBER Working Papers 28832, National Bureau of Economic Research, Inc.
- Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., and Zhuravskaya, E. (2015). Radio and the Rise of The Nazis in Prewar Germany *. *The Quarterly Journal of Economics*, 130(4):1885–1939.
- Ahiska, M. (2010). *Occidentalism in Turkey: Questions of Modernity and National Identity in Turkish Radio Broadcasting*. I.B. Tauris, London, 1 edition.
- Ahlquist, J. S. and Levi, M. (2011). Leadership: What It Means, What It Does, and What We Want to Know About It. *Annual review of political science*, 14(1):1–24.
- Akerlof, G. A. and Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics*, 115(3):715–753.
- Akerlof, R. and Holden, R. (2016). Movers and Shakers. *The Quarterly journal of economics*, 131(4):1849–.
- Akgüngör, S., Aldemir, C., Kuştepelı, Y., Gülcan, Y., and Tecim, V. (2011). The Effect of Railway Expansion on Population in Turkey, 1856-2000. *The Journal of Interdisciplinary History*, 42(1):135–157.
- Alesina, A., Baqir, R., and Easterly, W. (1999). Public Goods and Ethnic Divisions*. *The Quarterly Journal of Economics*, 114(4):1243–1284.
- Alesina, A. and Giuliano, P. (2015). Culture and Institutions. *Journal of Economic Literature*, 53(4):898–944.

- Alesina, A., Giuliano, P., and Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics*, 128(2):469–530.
- Alesina, A., Giuliano, P., and Reich, B. (2021). Nation-Building and Education. *The Economic Journal*, 131(638):2273–2303.
- Alesina, A. and Reich, B. (2015). Nation-building.
- Alvaredo, F., Assouad, L., and Piketty, T. (2019). Measuring Inequality in the middle east 1990–2016: The world’s most unequal region? *Review of Income and Wealth*, 65(4):685–711.
- Alvaredo, F., Atkinson, A., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2016). Distributional national accounts (dina) guidelines: Concepts and methods used in wid.world.
- Alvaredo, F., Piketty, T., Saez, E., Chancel, L., and Zucman, G. (2018). World inequality report 2018. Post-print, HAL.
- Andersen, J. J., Johannesen, N., Lassen, D. D., and Paltseva, E. (2017). Petro Rents, Political Institutions, and Hidden Wealth: Evidence from Offshore Bank Accounts. *Journal of the European Economic Association*, 15(4):818–860.
- Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Aslan, S. (2009). Incoherent State: The Controversy over Kurdish Naming in Turkey. *European journal of Turkish studies*.
- Aslan, S. (2011). Everyday Forms of State Power and the Kurds in the Early Turkish Republic. *International journal of Middle East studies*, 43(1):75–93.
- Assaad, R., Krafft, C., Roemer, J., and Salehi-Isfahani, D. (2018). Inequality of opportunity in wages and consumption in egypt. *Review of Income and Wealth*, 64(s1):S26–S54.
- Assouad, L. (2021). Rethinking the Lebanese Economic Miracle: The Extreme Concentration of Income and Wealth in Lebanon, 2005-2014. World Inequality Lab Working Papers halshs-02795688, HAL.
- Assouad, L., Chancel, L., and Morgan, M. (2018). Extreme inequality: Evidence from brazil, india, the middle east, and south africa. *AEA Papers and Proceedings*, 108:119–23.
- Atabaki, T. and Zürcher, E. J. (2004). *Men of Order : Authoritarian Modernization under Atatürk and Reza Shah*. I.B. Tauris, London ;
- Atkinson, A. and Piketty, T., editors (2007). *Top Incomes Over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*. Oxford University Press.

- Atkinson, A. and Piketty, T., editors (2010). *Top Incomes: A Global Perspective*. Oxford University Press.
- Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71.
- Aymes, M., Gourisse, B., and Massicard, E. (2014). *L'art de l'État en Turquie*. Meydan. Karthala.
- Aymes, M., Gourisse, B., and Massicard, E. (2015). *Order and Compromise: Government Practices in Turkey from the Late Ottoman Empire to the Early 21st Century*. Brill, Leiden, The Netherlands.
- Aytürk, I. (2004). Turkish Linguists against the West: The Origins of Linguistic Nationalism in Atatürk's Turkey. *Middle Eastern Studies*, 40(6):1–25.
- Aytürk, I. (2008). The First Episode of Language Reform in Republican Turkey: The Language Council from 1926 to 1931. *Journal of the Royal Asiatic Society*, 18(3):275–293.
- Bandiera, O., Mohnen, M., Rasul, I., and Viarengo, M. (2019). Nation-building Through Compulsory Schooling during the Age of Mass Migration. *Economic Journal*, 129(617):62–109.
- Bassi, V. and Rasul, I. (2017). Persuasion: A Case Study of Papal Influences on Fertility-Related Beliefs and Behavior. *American Economic Journal: Applied Economics*, 9(4):250–302.
- Bau, N. (2021). Can policy change culture? government pension plans and traditional kinship practices. *American Economic Review*, 111(6):1880–1917.
- Baumann, H., editor (2017). *Citizen Hariri: Lebanon's Neo-Liberal Reconstruction*. Oxford University Press.
- Bazzi, S., Fiszbein, M., and Gebresilasse, M. (2020). Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States. *Econometrica*, 88(6):2329–2368.
- Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2019). Unity in Diversity? How Intergroup Contact Can Foster Nation Building. *American Economic Review*, 109(11):3978–4025.
- Bazzi, S., Koehler-Derrick, G., and Marx, B. (2018). The Institutional Foundations of Religious Politics: Evidence from Indonesia. Sciences Po publications 25151, Sciences Po.
- Beck Knudsen, A. S. (2019). Those Who Stayed: Individualism, Self-Selection and Cultural Change during the Age of Mass Migration. Discussion Papers 19-01, University of Copenhagen. Department of Economics.

- Becker, S. O., Hsiao, Y., Pfaff, S., and Rubin, J. (2020). Multiplex Network Ties and the Spatial Diffusion of Radical Innovations: Martin Luther's Leadership in the Early Reformation. *American Sociological Review*, 85(5):857–894.
- Berkes, N. (1964). *The Development of Secularism in Turkey*. McGill University Press, Montreal.
- Bertrand, M. and Schoar, A. (2003). Managing with Style: The Effect of Managers on Firm Policies*. *The Quarterly Journal of Economics*, 118(4):1169–1208.
- Bibi, S. and Nabli, M. K. (2009). Income Inequality In The Arab Region: Data And Measurement, Patterns And Trends. *Middle East Development Journal (MEDJ)*, 1(02):275–314.
- Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Formation and persistence of oppositional identities. *European Economic Review*, 55(8):1046–1071.
- Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2016). Bend it like Beckham: Ethnic identity and integration. *European Economic Review*, 90(C):146–164.
- Bisin, A., Rubin, J., Seror, A., and Verdier, T. (2021). Culture, institutions and the long divergence. Working Paper 28488, National Bureau of Economic Research.
- Bisin, A. and Verdier, T. (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, 97(2):298–319.
- Bisin, A. and Verdier, T. (2017). On the Joint Evolution of Culture and Institutions. CEPR Discussion Papers 12000, C.E.P.R. Discussion Papers.
- Blanc, G. and Kubo, M. (2021). Schools, Language, and Nations: Evidence From a Natural Experiment in France.
- Blanchet, T., Fournier, J., and Piketty, T. (2022). Generalized pareto curves: Theory and applications. *Review of Income and Wealth*, 68(1):263–288.
- Blaydes, L. (2017). State building in the middle east. *Annual Review of Political Science*, 20(1):487–504.
- Blaydes, L. and Chaney, E. (2013). The feudal revolution and europe's rise: Political divergence of the christian west and the muslim world before 1500 ce. *American Political Science Review*, 107(1):16–34.
- Blouin, A. and Mukand, S. W. (2019). Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda. *Journal of Political Economy*, 127(3):1008–1062.
- Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting Event Study Designs: Robust and Efficient Estimation. Papers 2108.12419, arXiv.org.
- Bourguignon, F. and Morrisson, C. (2002). Inequality among world citizens: 1820-1992. *American Economic Review*, 92(4):727–744.

- Brockett, G. D. (2011). *How happy to call oneself a Turk: Provincial newspapers and the negotiation of a Muslim national identity*. University of Texas Press.
- Bulliet, R. W. (1978). First Names and Political Change in Modern Turkey. *International Journal of Middle East Studies*, 9(4):489–495.
- Burkhauser, R., Hérault, N., Jenkins, S., and Wilkins, R. (2016). What has been happening to UK income inequality since the mid-1990s? answers from reconciled and combined household survey and tax return data. NBER Working Papers 21991, National Bureau of Economic Research, Inc.
- Caesmann, M., Caprettini, B., Voth, H.-J., and Yanagizawa-Drott, D. (2021). Going Viral: Propaganda, Persuasion and Polarization in 1932 Hamburg. CEPR Discussion Papers 16356, C.E.P.R. Discussion Papers.
- Cagé, J., Dagherret, A., Grosjean, P., and Jha, S. (2020). Heroes and Villains: The Effects of Combat Heroism on Autocratic Values and Nazi Collaboration in France. CEPR Discussion Papers 15613, C.E.P.R. Discussion Papers.
- Callaway, B. and Sant'Anna, P. H. C. (2020). Difference-in-Differences with Multiple Time Periods.
- Cammett, M., editor (2014). *Compassionate communalism: welfare and sectarianism in Lebanon*. Cornell University Press.
- Cammett, M. and Diwan, I. (2019). *The Middle East*, chapter The Political Economy of Development in the Middle East. Singapore: CQ Press.
- Cammett, M., Diwan, I., Richards, A., and Waterbury, J. (2015). *A Political Economy of the Middle East*. Routledge.
- Cantoni, D., Chen, Y., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2017). Curriculum and Ideology. *Journal of Political Economy*, 125(2):338–392.
- Carvalho, J.-P. (2013). Veiling. *The Quarterly Journal of Economics*, 128(1):337–370.
- Carvalho, J.-P. and Sacks, M. (2021). Radicalization.
- Caymaz, B. and Szurek, E. (2007). La révolution au pied de la lettre. L'invention de "l'Alphabet turc" (The revolution at face value. The invention of the "Turkish alphabet"). *European journal of Turkish studies*, 6(6).
- Çagaptay, S. (2004). Race, Assimilation and Kemalism: Turkish Nationalism and the Minorities in the 1930s. *Middle Eastern Studies*, 40(3):86–101.
- Çagaptay, S. (2005). *Islam, Secularism and Nationalism in Modern Turkey: Who is a Turk?* Routledge.
- Cermeño, A. L., Enflo, K., and Lindvall, J. (2021). Railroads and Reform: How Trains Strengthened the Nation State. *British Journal of Political Science*, pages 1–21.

- Chaaban, J. (2016). I've got the power: Mapping connections between lebanon's banking sector and the ruling class. Working Papers 1059, Economic Research Forum.
- Chaney, E. (2016). Religion and the rise and fall of islamic science. Technical report.
- Colella, F., Lalive, R., Sakallı, S., and Thoenig, M. (2020). ACREG: Stata module to perform Arbitrary Correlation Regression. Statistical Software Components, Boston College Department of Economics.
- Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.
- Corm, G. (2005). *Le Liban contemporain: histoire et société*. Découverte/Poche: Essais, poche edition.
- Coulmont, B. (2014). *Sociologie des Prénoms*. La Découverte.
- Daher, A. (2014). *Le Hezbollah: Mobilisation et pouvoir*, chapter Chapitre IV: Le Hezbollah, un entrepreneur social?: Action sociale et mobilisation, ou l'édification de la "société de résistance", pages 147–188. Presses Universitaires de France.
- de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–96.
- Dell, M. and Querubin, P. (2017). Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies*. *The Quarterly Journal of Economics*, 133(2):701–764.
- DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M., and Zhuravskaya, E. (2014). Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia. *American Economic Journal: Applied Economics*, 6(3):103–32.
- DellaVigna, S. and Gentzkow, M. (2010). Persuasion: Empirical Evidence. *Annual Review of Economics*, 2(1):643–669.
- DellaVigna, S. and Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting*. *The Quarterly Journal of Economics*, 122(3):1187–1234.
- Depetris-Chauvin, E., Durante, R., and Campante, F. (2020). Building Nations through Shared Experiences: Evidence from African Football. *American Economic Review*, 110(5):1572–1602.
- Derenoncourt, E. (2018). 20. *The Historical Origins of Global Inequality*, pages 491–511. Harvard University Press.
- Devarajan, S. and Ianchovichina, E. (2018). A broken social contract, not high inequality, led to the arab spring. *Review of Income and Wealth*, 64(s1):S5–S25.
- Dippel, C. and Heblich, S. (2021). Leadership in Social Movements: Evidence from the "Forty-Eighters" in the Civil War. *American Economic Review*, 111(2):472–505.

- Diwan, I. and Haidar, J. I. (2021). Political connections reduce job creation: Firm-level evidence from Lebanon. *The Journal of Development Studies*, 57(8):1373–1396.
- Diwan, I., Malik, A., and Atiyas, I., editors (2019). *Crony Capitalism in the Middle East, Business and Politics from Liberalization to the Arab Spring*. Oxford University Press.
- Dowrick, S. and Akmal, M. (2005). Contradictory trends in global income inequality: A tale of two biases. *Review of Income and Wealth*, 51(2):201–229.
- El-Katiri, L., Fattouh, B., and Segal, P. (2011). Anatomy of an oil-based welfare state: Rent distribution in Kuwait. *Kuwait Program on Development, Governance, and Globalization in the Gulf States*.
- Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). Media and Political Persuasion: Evidence from Russia. *The American Economic Review*, 101(7):3253–3285.
- Fabbe, K. (2019). *Disciples of the State? : Religion and State-Building in the Former Ottoman World*. Cambridge University Press.
- Fearon, J. D. and Laitin, D. D. (2003). Ethnicity, Insurgency, and Civil War. *The American Political Science Review*, 97(1):75–90.
- Fernandez, R. (2010). Does Culture Matter? NBER Working Papers 16277, National Bureau of Economic Research, Inc.
- Fouka, V. (2019). Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. *The Review of Economic Studies*, 87(1):204–239.
- Garbinti, B., Goupille-Lebret, J., and Piketty, T. (2018). In France, 1900–2014: Evidence from distributional national accounts (DINA). *Journal of Public Economics*, 162:63–77. In Honor of Sir Tony Atkinson (1944–2017).
- Gaspard, T. (2004). A Political Economy of Lebanon, 1948–2002: The Limits of Laissez-Faire. *International Journal of Middle East Studies*, 37(4):640–642.
- Gerber, A. S. and Green, D. P. (2000). The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment. *The American Political Science Review*, 94(3):653–663.
- Ginsburgh, V. and Weber, S. (2020). The Economics of Language. *Journal of Economic Literature*, 58(2):348–404.
- Giuliano, P. and Nunn, N. (2021). Understanding Cultural Persistence and Change. *Review of Economic Studies*, 88(4):1541–1581.
- Gorodnichenko, Y. and Roland, G. (2011). Which Dimensions of Culture Matter for Long-Run Growth? *The American Economic Review*, 101(3):492–498.

- Grosfeld, I., Rodnyansky, A., and Zhuravskaya, E. (2013). Persistent Antimarket Culture: A Legacy of the Pale of Settlement after the Holocaust. *American Economic Journal: Economic Policy*, 5(3):189–226.
- Grosfeld, I. and Zhuravskaya, E. (2015). Cultural vs. economic legacies of empires: Evidence from the partition of Poland. *Journal of Comparative Economics*, 43(1):55–75.
- Gürpınar, D. (2012). What is in a Name? The Rise of Turkic Personal Male Names in Turkey (1908–38). *Middle Eastern Studies*, 48(5):689–706.
- Habyarimana, J., Humphreys, M., Posner, D. N., and Weinstein, J. M. (2009). *Coethnicity: Diversity and the Dilemmas of Collective Action*. Russell Sage Foundation Press.
- Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, 20(1):25–46.
- Hariri, J. G. (2015). A contribution to the understanding of middle eastern and muslim exceptionalism. *The Journal of Politics*, 77(2):477–490.
- Hassine, N. B. (2015). Economic Inequality in the Arab Region. *World Development*, 66(C):532–556.
- Hermalin, B. (1998). Toward an Economic Theory of Leadership: Leading by Example. *American Economic Review*, 88(5):1188–1206.
- Hertog, S. (2017). The political economy of distribution in the middle east: Is there scope for a new social contract? *Revue internationale de politique de développement*, 7.
- Hlasny, V. and Verme, P. (2018). Top Incomes and the Measurement of Inequality in Egypt. *World Bank Economic Review*, 32(2):428–455.
- Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences*, 107(35):15367–15372.
- Human Rights Watch, H. (2013). South Asia: Protect Migrant Workers to Gulf Countries.
- Jawad, R. (2009). *Social Welfare and Religion in the Middle East: A Lebanese Perspective*. Policy Press.
- Jones, B. F. and Olken, B. A. (2005). Do Leaders Matter? National Leadership and Growth Since World War II. *The Quarterly Journal of Economics*, 120(3):835–864.
- Jordá, V. and Niño-Zarazúa, M. (2019). Global inequality: How large is the effect of top incomes? *World Development*, 123:104593.
- Kafadar, C. (1995). *Between Two Worlds : The Construction of the Ottoman State*. University of California Press, Berkeley, CA.

- Kapiszewski, A. (2006). Arab versus asian migrant workers in the gcc countries*.
- Kelly, M. (2019). The Standard Errors of Persistence. Working Papers 201913, School of Economics, University College Dublin.
- Kinninmont, J. (2015). Future Trends in the Gulf.
- Kocatürk, U. (1988). *Doğumundan ölümüne kadar kaynakçalı Atatürk günlüğü*. Ankara: Türkiye İş Bankası Kültür Yayınları, 3rd ed. edition.
- Kuran, T. (1995). *Private truths, public lies : the social consequences of preference falsification*. Harvard University Press, Cambridge, Mass.
- Kuran, T. (2004). Why the middle east is economically underdeveloped: Historical mechanisms of institutional stagnation. *Journal of Economic Perspectives*, 18(3):71–90.
- Kuran, T. (2010). *The Long Divergence: How Islamic Law Held Back the Middle East*. NJ: Princeton University Press.
- Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 45(1):1–28.
- Laithy, H., Abu-Ismaïl, K., and Hamdan, K. (2008). Poverty, Growth and Income Distribution in Lebanon.
- Lakner, C. and Milanovic, B. (2013). Global income distribution : from the fall of the Berlin Wall to the great recession. Policy Research Working Paper Series 6719, The World Bank.
- Lamprou, A. (2015). *Nation-Building in Modern Turkey: The "People's Houses", the State and the Citizen*. I.B. Tauris., London and New York.
- Lewis, G. (1974). *Modern Turkey*. Nations of the modern world. Praeger, New York, [4th ed.] edition.
- Lewis, G. (1999). *The Turkish language reform, a catastrophic success*. New York: Oxford University Press.
- Lijphart, A. (1969). Consociational democracy. *World Politics*, 21(2):207–225.
- Loeper, A., Steiner, J., and Stewart, C. (2014). Influential Opinion Leaders. *The Economic Journal*, 124(581):1147–1167.
- Mango, A. (1999). *Atatürk*. John Murray London.
- Mardin, Ş. (2002). Playing Games with Names. In *Fragments of Culture : the Everyday of Modern Turkey*, edited by Deniz Kandiyoti and Ayşe Saktanber., pages 143–156. I.B. Tauris.
- Mateos, P. (2013). *Names, Ethnicity and Populations* . Heidelberg, Germany: Springer.

- Metin, H. (1998). *İsmet İnönü: The Making of a Turkish Statesman*. Social, Economic and Political Studies of the Middle East and Asia. Brill, Leiden.
- Miguel, E. (2004). Tribe or nation? nation building and public goods in kenya versus tanzania. *World Politics*, 56(3):327–362.
- Milanovic, B. (2002). True world income distribution, 1988 and 1993: First calculation based on household surveys alone. *The Economic Journal*, 112(476):51–92.
- Ministry of Planning (1960). *Needs and developing possibilities in Lebanon: preliminary study*. Besoins et possibilités de développement du Liban: étude préliminaire. Mission Irfed.
- Morgan, M. (2017). Extreme and Persistent Inequality : New Evidence for Brazil Combining National Accounts , Surveys and Fiscal Data , 2001-2015.
- Murphy, K. M. and Shleifer, A. (2004). Persuasion in Politics. *American Economic Review*, 94(2):435–439.
- Mylonas, H. and Tudor, M. (2021). Nationalism: What we know and what we still need to know. *Annual Review of Political Science*, 24(1):109–132.
- Ncube, M. and Anyanwu, J. C. (2012). Inequality And Arab Spring Revolutions In North Africa And The Middle East. *Africa Economic Brief*, 3(7).
- Nişanyan, S. (2010). *Adımı Unutan Ülke: Türkiye’de Adı Değiştirilen Yerler Sözlüğü*. Charles H. Herr and Co Chicago.
- Novokmet, F., Piketty, T., and Zucman, G. (2017). From Soviets to Oligarchs: Inequality and Property in Russia 1905-2016.
- Ottinger, S. and Voigtländer, N. (2020). History’s Masters: The Effect of European Monarchs on State Performance. NBER Working Papers 28297, National Bureau of Economic Research, Inc.
- Özdemir, N. (2006). *Cumhuriyet Dönemi Karayolu Politikası (1932-1960)*. PhD thesis. Doktora Tezi, T.C. Ankara Üniversitesi, Sosyal Bilimler Enstitüsü Tarih.
- Pamuk, Ş. (2018). *Uneven Centuries: Economic Development of Turkey since 1820*, volume 75 of *The Princeton Economic History of the Western World*. Princeton University Press, Princeton.
- Parreira, C. (2020). *The Art of Not Governing: Local Politics in Postwar Lebanon*. PhD thesis, Stanford University.
- Piketty, T., Saez, E., and Zucman, G. (2016). Distributional National Accounts: Methods and Estimates for the United States.
- Piketty, T., Yang, L., and Zucman, G. (2017). Capital Accumulation , Private Property and Rising Inequality in China , 1978-2015.

- Rijkers, B., Freund, C., and Nucifora, A. (2017). All in the family: State capture in Tunisia. *Journal of Development Economics*, 124(C):41–59.
- Rubin, J. (2017). *Rulers, Religion, and Riches: Why the West Got Rich and the Middle East Did Not*. Cambridge Studies in Economics, Choice, and Society. Cambridge University Press.
- Saavedra, M. (2021). Kenji or Kenneth? Pearl Harbor and Japanese-American assimilation. *Journal of Economic Behavior and Organization*, 185:602–624.
- Saez, E. and Zucman, G. (2016). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data *. *The Quarterly Journal of Economics*, 131(2):519–578.
- Said, M. (2007). The Fall and Rise of Earnings and Inequality in Egypt: New Evidence From the ELMPS, 2006. Working Papers 708, Economic Research Forum.
- Sakallı, S. O. (2019). Secularization and Religious Backlash: Evidence from Turkey.
- Saleh, M. (2017). A ‘new’ economic history of the Middle East and North Africa (MENA) region. *The Economics of Transition*, 25(2):149–163.
- Satyanath, S., Voigtländer, N., and Voth, H.-J. (2017). Bowling for Fascism: Social Capital and the Rise of the Nazi Party. *Journal of Political Economy*, 125(2):478 – 526.
- Scott, J. C., Tehranian, J., and Mathias, J. (2002). The Production of Legal Identities Proper to States: The Case of the Permanent Family Surname. *Comparative Studies in Society and History*, 44(1):4–44.
- Spears, I. (2002). Africa: The limits of power-sharing. *Journal of Democracy*, 13(3).
- Spencer, R. F. (1961). The Social Context of Modern Turkish Names. *Southwestern Journal of Anthropology*, 17(3):205–218.
- Spolaore, E. and Wacziarg, R. (2013). How Deep Are the Roots of Economic Development? *Journal of Economic Literature*, 51(2):325–69.
- Szurek, E. (2013). *Gouverner par les mots : une histoire linguistique de la Turquie nationaliste*. PhD thesis. Thèse de doctorat dirigée par Georgeon, François Histoire et civilisations Paris, EHESS 2013.
- Szurek, E. (2015). The Linguist and the Politician: the Türk Dil Kurumu and the Field of Power in the 1930-40s. pages 68 – 96. Brill, Leiden, The Netherlands.
- Szurek, E. (2018). *Le Moyen-Orient. Fin XIXe-XXe siècle*, chapter Penser la révolution turque, pages 365–400. Seuil.
- Szurek, E. (2020). “Yan, Of, Ef, Viç, İç, İs, Dis, Pulos...” : the Surname Reform, the “Non-Muslims”, and the Politics of Uncertainty in Post-genocidal Turkey. In *Arabic and its Alternatives, Religious Minorities and their Languages in the Emerging Nation States of the Middle East (1920-1950)*, pages 77–110. Brill.

- Tilly, C. (1993). *Coercion, Capital and European States, A.D. 990-1992*. Wiley-Blackwell.
- Traboulsi, F. (2012). Social Classes and Political Power in Lebanon. *Heinrich-Boll Stiftung Middle East*, pages 1–133.
- Tuna, M. (2018). The Missing Turkish Revolution: Comparing Village-level Change and Continuity in Republic Turkey and Soviet Central Asia, 1920-50. *International Journal of Middle East Studies*, 50(1):23–43.
- Türköz, M. (2018). *Naming and Nation-building in Turkey: The 1934 Surname Law*. Palgrave Macmillan US, New York, NY, U.S.A.
- Ülker, E. (2008). Assimilation of the Muslim communities in the first decade of the Turkish Republic (1923-1934). *European Journal of Turkish Studies. Social Sciences on Contemporary Turkey*.
- UNDP (2002). *Globalization: Towards a Lebanese Agenda*.
- van der Weide, R., Lakner, C., and Ianchovichina, E. (2016). Is inequality underestimated in egypt? evidence from house prices.
- Verdeil, É. (2006). Reconstruire encore. . . Le bilan des destructions de la guerre de 2006 au Liban. In *Une guerre au Liban: été 2006*, pages 1–11. Cahiers Libres, la découverte edition.
- Verdier, T. and Zénou, Y. (2018). Cultural Leader and the Dynamics of Assimilation. *Journal of Economic Theory*, 175(C):374–414.
- Voigtlaender, N. and Voth, H.-J. (2014). Highway to Hitler. Working Paper 20150, National Bureau of Economic Research.
- Voigtländer, N. and Voth, H.-J. (2012). Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany*. *The Quarterly Journal of Economics*, 127(3):1339–1392.
- Wahba, J. (1996). Earnings and regional inequality in egypt. Working Papers 9613, Economic Research Forum.
- Wang, T. (2021). Media, Pulpit, and Populist Persuasion: Evidence from Father Coughlin. *American Economic Review*, 111(9):3064–92.
- Weber, E. (1976). *Peasants into Frenchmen : the modernization of rural France, 1870-1914* . Stanford University Press Stanford, Calif.
- Weber, M. (1921). *Economy and Society: A New Translation*. New edition: Cambridge: Harvard University Press, 2019, ed. and trans. Keith Tribe edition.
- Wimmer, A. and Feinstein, Y. (2010). The rise of the nation-state across the world, 1816 to 2001. *American Sociological Review*, 75(5):764–790.

- World Bank, W. (2016). Lebanon- Promoting poverty reduction and shared prosperity: a systematic country diagnostic.
- Yilmaz, H. (2013). *Becoming Turkish: Nationalist Reforms and Cultural Negotiations in Early Republican Turkey 1923-1945*. Syracuse University Press.
- Yonzan, N., Milanovic, B., Morelli, S., and Gornick, J. (2021). Drawing a Line Comparing the Estimation of Top Incomes Between Tax Data and Household Survey Data. CSEF Working Papers 600, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.
- Zeydanlıoğlu, W. (2012). Turkey's Kurdish Language Policy. *International Journal of the Sociology of Language*, 2012(217):99–125.
- Zucman, G. (2013). The missing wealth of nations: Are Europe and the U.S. net debtors or net creditors? *The Quarterly Journal of Economics*, 128(3):1321–1364.
- Zürcher, E. J. (2012). In the Name of the Father, the Teacher and the Hero: the Atatürk personality cult in Turkey. In *Political Leadership, Nations and Charisma*, edited by Vivian Ibrahim and Margit Wunsch, pages 143–156. Routledge.
- Zürcher, E. J. (2017). *Turkey : A Modern History*. Library of modern Turkey ; 27. I.B. Tauris & Co. Ltd, London ;, fourth edition. edition.

APPENDIX

A. APPENDIX TO CHAPTER 1

A.1 Data sources and Methodology

This section provides details on the data, method and assumptions made at each step of the estimation procedure, as already summarized in Section 1.2 of the main paper. An online appendix that includes all raw data sources and computer codes is available at <https://wid.world/country/lebanon/>.

A.1.1 Step 1: Household Survey Series

The first step consists in generating a first "raw" income distributions using survey data. There are three nationally representative surveys in Lebanon, in 1997, 2004 and 2007. As discussed in the paper, I could not access data for the 1997 survey, either in the form of tabulations by range of income, or in the form of micro data. For the 2004 and 2007 surveys, micro-data are also inaccessible. The Lebanese Central Administration of Statistics however publishes survey reports including tabulations for the 2004 and 2007 survey. The raw tabulations are available in the directory "/HouseholdSurveyData/" in the data files online.

I use two tables from the 2007 survey reports, giving the number of households in 13 income groups "before" and "after the Israeli 2006 war" (2005 and 2007). The DINA guidelines recommends to use the same unit of observation across country: the "adult" individual (aged 20 and more). To express income in terms of adults, I assume income is equally split between adult household members, that is I divide household income by the number of adults in each household. As no additional information is available, I apply households have the same size across income group, and I take the average adults/children ratio in the country: if high earners have fewer children than average, inequality is slightly underestimated. I then apply the generalized Pareto interpolation techniques developed by Blanchet, Fournier and Piketty (2017) to both tables, to estimate the full distribution of income expressed in generalized percentiles (or g-percentiles) between for 2005 and 2007.

The generalized Pareto interpolation technique can be applied to tabulations providing three pieces of information: income thresholds, household frequencies and the average income per group. Unfortunately, the latter was not available in the Lebanese tabulations. To perform the estimation, I need to make an assumption on the form of the tail of the distribution at the top. In the benchmark estimates, I assume that the last group (approximately the top 0.5 percent in both tables) is characterized by an inverted Pareto coefficient of 2. This assumption has no impact on the final series (this is why I do not display them). Most importantly, given that the top will be corrected with the fiscal data, this assumption has little effect. I nevertheless highlighted this step in the decision tree (Figure 1.11).

Finally, I simply upgrade the 2007 distribution by the ratio of per adult national income of a given year between 2008 and 2014 (divide by the average per adult national income and multiply by the average per adult national income in a given year), and get the 2005-2014 survey series. I use the 2005 distribution for the years 2005 and 2006, in a similar way. By definition, inequality is constant between 2005 and 2006 and between 2007-2014. Figure A.1 shows the effect of using only the 2007 survey versus the 2004 and

2007 survey.

A.1.2 Step 2: Fiscal Series

To estimate the fiscal series, I need to make two additional assumptions: the first concerns the definition of income; the second concerns the choice of correction profile to link the fiscal data to the survey distribution.

Definition of income

As explained in the DINA guidelines, it is critical to be precise about the income concepts when combining survey, fiscal data and national accounts. Unfortunately, the survey data do not enable me to precisely define income. As for the fiscal data, they only provide information on the "taxable income" (i.e. income subject to income tax, after a number of deductions allowed by the tax legislation). The deductions are very extensive in the Lebanese case (see Section A.2.2 below). In particular, there are large lump-sum deductions for professional expenses of self-employment income. Additionally, taxable income, from which benefits and allowances are deducted, is significantly smaller than the fiscal income, defined as the sum of all income items legally subject to taxation, before any deduction (Alvaredo et al. (2016). I therefore assume for my benchmark estimates that the ratio between taxable income and fiscal income is equal to $r=80$ percent. I also estimate as robustness checks a number of variant using other ratios ($r=70$ and $r=90$ percent). Figure A.2 shows the impact of the choice of the income definition on the final series. This assumption has a relatively sizable effect on the final series, mainly due to the small share of survey income in total national income. This is why I take a relatively conservative hypothesis (80 percent) given the deductions allowed in the Lebanese tax law.

Correction profiles

To correct the top of the survey distribution with fiscal data, I first need to make an assumption on the "ratio" of the national income distribution each source can reliably cover. Then, I need to make an assumption on how to link both sources. My benchmark correction is based upon the following assumption: the survey data is reliable below percentile $p_1 = 0.8$, the fiscal data is reliable above $p_2 = 0.99$ and I assume that the quantile ratio upgrade factor $f(p)$ rises piecewise-linearly from $f(p_1) = 1$ to the observed fiscal/survey ratio $f(p_2)$ between p_1 and p_2 , with a small and rising slope between $p_1 = 0.8$ and $p=0.9$ and a constant linear slope between $p=0.9$ and $p_2 = 0.99$. I also consider other profiles: one where I assume the survey data to be reliable below percentile $p_1 = 0.9$, the fiscal data to be reliable above $p_2 = 0.99$, and a linear profile of $f(p)$ between p_1 and p_2 (profile 2). In other profiles, I assume a concave (declining slope) and a convex (increasing slope) of $f(p)$ between p_1 and p_2 (profile 3 and 4). As shown in Figure A.3, the variants have also a non-negligible impact on the results, especially for the top 1 percent.¹ In any case, two remarks should be made: (1) the fiscal correction is the largest

1. For the detailed definition of the four profiles and the corresponding factors, see the excel file "CompCorrectionCoeffLeb.xlsx", sheet "CompUpFactorLeb", in the online appendix. Unsurprisingly, the more

in magnitude and leads to a large upward correction of the survey-based distributions (see Figure A.4 for the decomposition of the effect of each correction) (2) the variants on the fiscal correction can have a strong effect.²

A.1.3 *Missing Capital income*

Finally, I proceed to the last correction, which attempts to account for non-reported and tax-exempt capital income. There are two steps.

Estimating the amount of missing capital income

The first step consists in estimating the amount of missing capital income. Here again, important differences with Piketty, Yang and Zucman (2017) and Novokmet, Piketty and Zucman (2017), which use a similar methodology for China and Russia, should be noted. First, the amount of capital income absent from the Lebanese fiscal data are significantly higher. My correction needs to account for both tax-exempt and movable capital income, which are taxed but not reported in my datasource. Hence the denomination "missing" capital income as opposed to solely "non-fiscal". Second, national accounts are not disaggregated enough to estimate the missing amounts and do not display detailed enough subcomponents of national income. Only the generation and allocation of primary income accounts of the national economy (S1) are displayed, without details for the different sectors. The only sub-sectors present are the general government (S13) and Banks (S122). The amounts recorded are themselves not disaggregated enough to identify each income source. This is why I complement them with government reports on tax revenues and recover proxies for the amounts of income missing, by dividing the amount of taxes collected by the corresponding tax rates applied in the law. The idea is to recover the amounts of capital income generated in the economy by dividing the amount of taxes collected by the tax rates defined in the fiscal law. More precisely, I derive from the government reports: (1) the amount of capital gains and dividends accruing to the households, and taxed at flat tax rate of 10 percent under the third title of the personal income tax law (2) the amount of interest income received in the private sector in the total economy and hit by a rate of 5 percent (3) the imputed rents from housing taxed at 4 percent (4) undistributed profits of privately owned corporations. I find that they respectively represent 3, 8, 3 and 8 percent of national income on average over the period.³ While I find that the missing income should represent approximately 22 percent

the rising part of the $f(p)$ profile is pushed toward p_2 , the smaller the total upgrade to the top 10 percent share; and the more the rising part of $f(p)$ is pushed toward p_1 , the larger the total upgrade to the top 10 percent share. As long as there is no income tax data covering the entire top 10 percent, there is no way to be sure about this.

2. Recent research by Yonzan et al. (2020) suggests that the appropriate cutoff might be 0.90, using data from the US, France and Germany. This corresponds to my profile 2. Given the impact of this hypothesis on the final estimates, I chose a more conservative hypothesis.

3. Raw data from the reports and all computations can be found in the file "EstimatingMissingCapitalIncome.xlsx", in the directory GpinterIncome. In particular, as the total taxes on profits (taxed under the PIT and the corporate tax) are put together, I subtract from the aggregate tax revenues from profits the total amount reported in the fiscal data and divide by the corporate tax rate (of 15 percent) the remaining amount to get the total amount of profits subject to the corporate tax. Likewise, the taxes on property and

of national income. If we recall Figure 1.2, this means that there remains uncertainty for 8 percent of total national income that are re-allocated proportionally. Table A.1 sums up incomes the total amounts of income that can be inferred from the fiscal data, the national accounts or the finance reports. It also displays the amounts left (that encompass tax evasion, deductions and exemptions and other non-fiscal income and finally income from the informal sector, in part captured by the survey data). Figure A.5 shows how the amount of missing capital income impacts the final estimates.

Estimating the joint distributions of fiscal and non-fiscal income

Next, in order to estimate the final distribution of total personal income (y_p), I need to make an assumption about the distribution of missing capital income (y_m) and about the structure of correlation between the fiscal income distribution (y_f) and the missing income distribution (y_m). Regarding the distribution of (y_m), I assume it follows the same distribution as the distribution of wealth which is standard in the literature.⁴ Finally, I apply a proportional upgrade factor to transform the distribution of personal income (y_p) = (y_f) + (y_m) into the distribution of national income y . By construction this has no impact on income shares (the objective is to make income levels comparable across countries and over time). Regarding the correlation structure between (y_f) and (y_m), I use the family of Gumbel copulas, characterized by the following functional form:

$$F(u, v) = \exp[-((- \log u)^\theta + (- \log v)^\theta)^{\frac{1}{\theta}}] \quad (\text{A.1})$$

where $0 \leq u, v \leq 1$ are the ranks in the two distributions and $F(u, v)$ is the two-dimensional cumulative distribution, that is the fraction of the population with ranks below u in the first dimension and below v in the second dimension. If $\theta = 1$ then $F(u, v) = uv$, i.e. the two distributions are entirely independent. Conversely if $\theta = +\infty$ then both dimensions are perfectly correlated. On the basis of observed two-dimensional distributions in countries with high-quality fiscal data (such as the United States or France), it appears that the Gumbel parameters are typically in the 2.5-3.5 range. I use $\theta = 2$ for my benchmark estimates as a conservative assumption. The choice of the parameter has a relatively small impact on the final series (see Figure A.6 for sensitivity checks).

Wealth Series

As explained in the main paper, the methodology used to obtain the Lebanese wealth distribution is similar to the one used by Novokmet, Piketty and Zucman (2017) for Russia. The data sources available to estimate wealth inequality in Lebanon are very limited and at this stage I only have billionaire data. I proceed as follows. First, I compute average standardized distributions of wealth for the US, France and China from

on rental income are reported together, so I also remove first the total amount of taxes collected from build property revenues from my fiscal data and then apply the tax rate (of 4 percent) to the remaining amount.

4. Capital income and incomes which evade taxes, tend to more unequally distributed than labor income. See Section A.1.3 for the estimation of the wealth distribution

WID.world series (that is, I divide all thresholds and bracket averages for all 127 generalized percentiles by average wealth, and I compute the arithmetic average for the three countries). Variations across countries and over time in these standardized wealth distributions mostly happen above $p_0 = 0.99$. Below p_0 , the ratios of the different percentile thresholds to average wealth are relatively stable. Therefore I choose to use this average US-France-China normalized distribution for Lebanon below p_0 . Second, I need to determine the Lebanese total personal wealth per adult, so as to adapt this average US-France-China normalized distribution to Lebanon. Contrarily to the Russian case, there is for the moment no estimate of the total stock of personal wealth in Lebanon. I therefore take the average wealth/income ratios available in WID.world (which equals to 300 percent of national income), and apply it to the Lebanese national income. I hereby assume that (1) wealth is as concentrated in Lebanon as what is currently observable in other countries with adequate data and (2) that if, on average, countries own a stock of capital equals to 300 % of their national income, Lebanon owns as least as much. Finally, I use information on Lebanese billionaires to adjust the top of the distribution and to take into account the extremely high share of billionaires' wealth, compared to the total national income. The difficult question is to know how to link the distribution from p_0 to the billionaire level, and also to make an assumption about the average number n of adults per billionaire family (sometime Forbes includes very large family groups in the same billionaire family, sometime it is just one individual or one married couple). I first re-estimate 127 generalized percentile within the top 1 percent of the normalized distribution in order to reach billionaire level. In the benchmark series I assume $n=5$ and a linear correction factor $f(p)$ from $p_0=0.99$ up to billionaire level (because this seems to work relatively well for the US, France and China). Figure A.7 shows variant series based upon alternative assumptions for billionaires' family size: $n=2,4,6,8$ instead of $n=5$. The assumptions lead to relatively large differences in the wealth distribution (up to 2 percentage points). In any case, even the most conservative series lead to high wealth shares.

A.2 Lebanese Income Tax

A.2.1 Presentation of the Personal Income Tax Law

The Lebanese Income Tax was created in 1944 (Law 12/4/1944) and amended in 1959 (Decree-Law 144, 6/12/1959). The text of 1959 is still the basis of the current fiscal system. The 1959 income tax is a schedular, progressive and individual tax which taxes the different sources of income separately. It is divided into three main categories: a tax on profits from industrial, commercial and non-commercial activities levied according to a real or lump sum scheme (Title I), a tax on wages and salaries (Title II) and a tax on built property revenues (Title III). Next to the PIT, incomes from movable capital including interests and dividends are taxed according to a flat rate. This section draws extensively from Daher (2002).

1. **Title I: tax on profits from industrial, commercial and non-commercial activities:**
This concerns only the business income made by a sole proprietor (professional, individual company, individuals in Small corporation). These profits are taxed at

progressive rates between 4 percent and 21 percent. Business incomes made by a limited partnership (joint stock companies and limited liability companies are either put in reserve and serve for the company self-financing (in this case, they are not taxed) or they are distributed as interests or dividends to the partners (in this case they are not taxed under the PIT, but subject to a flat tax rates for revenues from moveable capital).

2. **Title II: Salaries and wages and pension salaries:** this tax concerns all types of labor income: wages and salaries, including bonuses, commissions, compensation, allowances, grants, benefits in cash and kind, overtime hours, pensions and annuities (Article 46 of D.L. no. 144/1959), after deductions of the allowances and charges. The tax is levied at source and declared annually by the employers, at progressive rates between 2 percent and 20 percent (personal income, reported in the database).
3. **Title III: Built property revenues:** It is charged on the flow of income generated by the ownership of a built property, according to a progressive tax scale (4-14 percent), on built property (personal income, reported in the database). There exist also a tax on built property, which is charged on the stock (4 percent of the value of the real-estate, non reported in the database).

A.2.2 Income definition and deductions

In this section, I present in further detail the variables reported in the fiscal database, by referring to the Lebanese Income tax Law and the 2010 tax forms.⁵ As explained in the main paper, three variables are reported for labor income, and business income:

1. Salaries and wages

- **The labor gross income**, which comprises the main salary/daily wages, representation remuneration, bonuses, commissions and overtime, family compensation for the spouse, family compensation for the children, allowances given to bear the expenses of the activity (transportation compensation, car allowance, residence allowance, food allowance, clothing allowance), fund compensations, health insurances of all types, educational grants, marriage grants, birth grants, assistance in case of illness, assistance in case of death, other grants and benefits.
- **The labor income subject to tax**, obtained after deducing from the gross labor income the compulsory social contributions, the allowances covering expenses linked with the professional activity and all the grants and benefits.⁶
- **Total amount of tax paid**

2. Self-employment income

- **Total turnover** made in a given year
- **The corresponding profit subject to tax**, equal to the turnover multiplied by a given rate in order to take into account charges and expenses endured during

5. See Figure A.8 for the general tax form.

6. Article 50, Law 144 (06/12/1959) modified by Laws 27 (07/19/1980), 7 (08/10/1985) and 89 (09/07/1991). See Tax form R6, Figure A.9.

- the activity.⁷ Self-employment incomes are taxed according to a lump-sum scheme. The rate applied varies between 3 percent and 65 percent depending on the activity.⁸ In the database, the effective coefficient applied is on average 30 percent for all years.
3. **Other business incomes.** For partners in partnerships and individuals in S-corporations:
 - **The actual total revenue**, defined as the turnover plus the overall financial and non-financial investment revenues.⁹
 - **The corresponding profit subject to tax**, which is equal to actual total revenue minus the expenses and costs incurred during the activity¹⁰, minus the exonerated incomes (grants and donations). The non-deductible revenues are capital interests, investments and expenses made to earn capital gains, taxes paid to a foreign government, losses incurred by branches settled abroad, representation remuneration distributed to employees and exceeding 10 percent of their wages
 4. **Built property revenues**, excluding persons living in their own dwelling: the taxable income after deduction and amount of tax paid are available.

7. The charges are "Sales of merchandise, consumption material, wages, salaries and other benefits, employees and wage-earners insurance, social security subscriptions, commissions paid to third parties, car and transportation expenses, banking commissions, interests and expenses, legal expenses, consultancies and similar expenses, maintenance and repair expenses, rent or investment, other office expenses, taxes, fees, and permits, accommodation, traveling expenses, promotion and advertisement, institution/profession activity insurance expenses, amortization" (see tax form F3, Figure A.10).

8. Decree 4169/1 (8/16/1993) modified by the Decree 5/1 (11/1/2000).

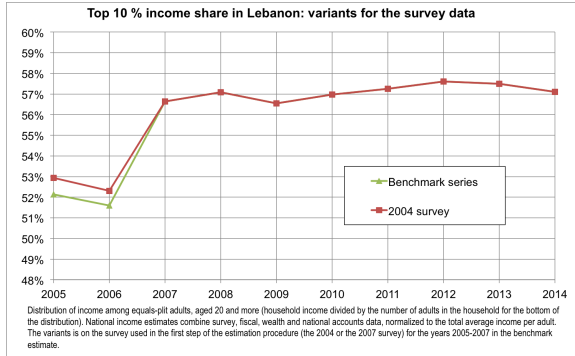
9. Common operations dividends, placement and participation bonds revenues, net profit from placement bonds wavering, revenues from other movables, similar interests and revenues, positive exchange rate differences, recoveries from financial provisions (tax forms F16-1 and F16-2, Figures A.11 and A.12).

10. The costs comprise: "the overall cost (sold merchandise, sold production, work and services provision cost), external services (royalties, rents etc.), employees charges (including social security contributions), tax fees and charges, the depreciation and investment provision allocations, interests on loans for the company's needs".

A.3 Additional Figures and Tables

Figure A.1 – Variants for the Step 1: Effect of the Survey Source

(a) Top 10%



(b) Top 1%

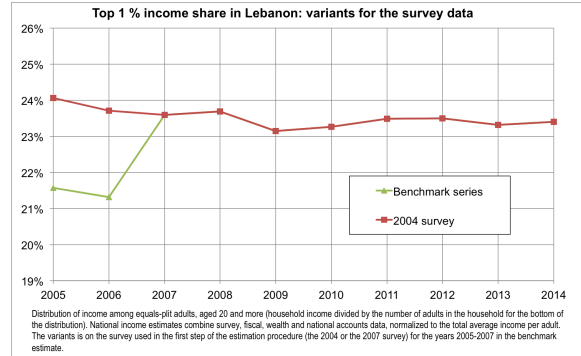
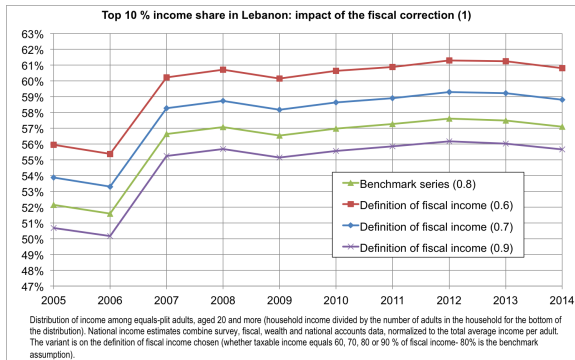


Figure A.2 – Variants for the Step 2: Effect of the Fiscal Income Definition

(a) Top 10%



(b) Top 1%

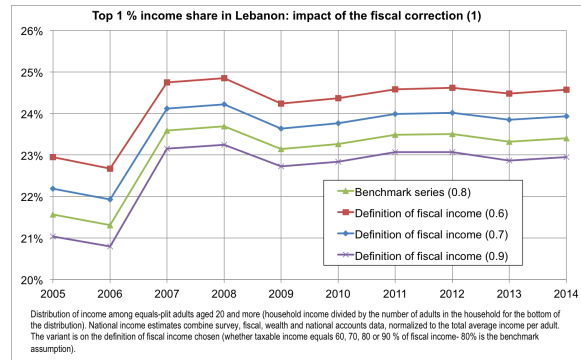


Figure A.3 – Variants for the Step 2: Effect of the Correction Profile

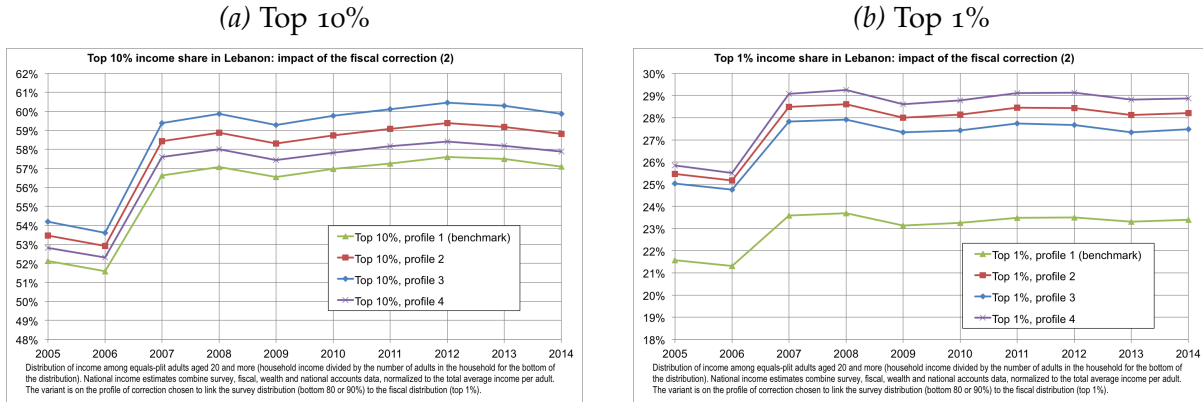


Figure A.4 – Impact of each correction on the Lebanese income share, 2005-2014

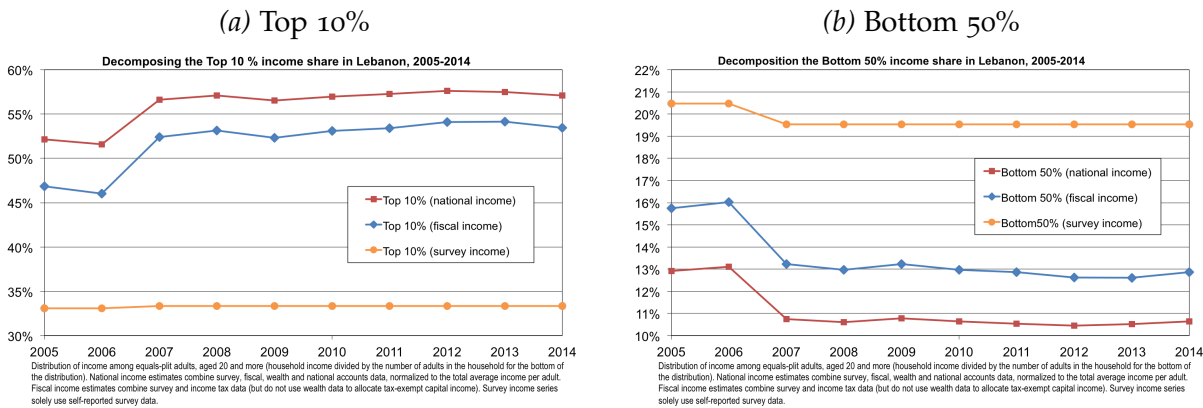


Figure A.5 – Variants for the Step 3: Effect of the size of the missing amount of re-allocate

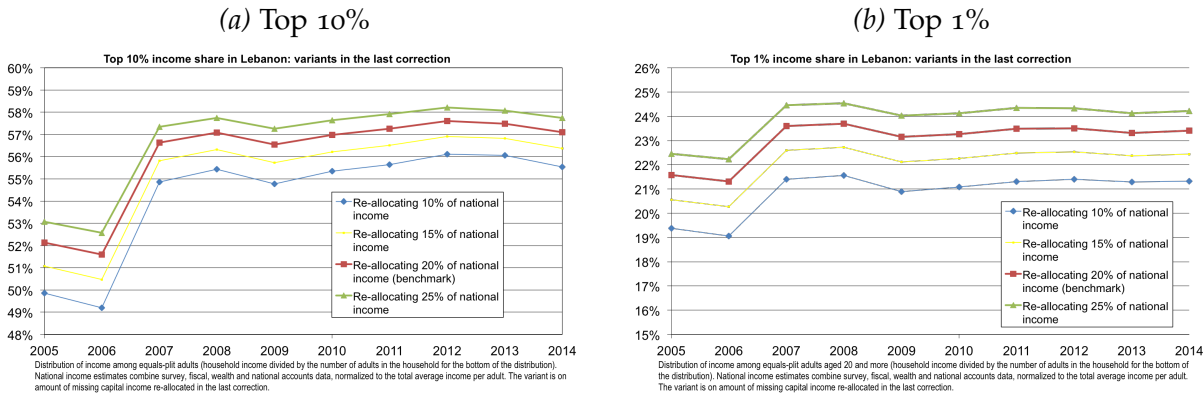


Figure A.6 – Variants for the Step 3: Effect of the size of the missing amount of re-allocate

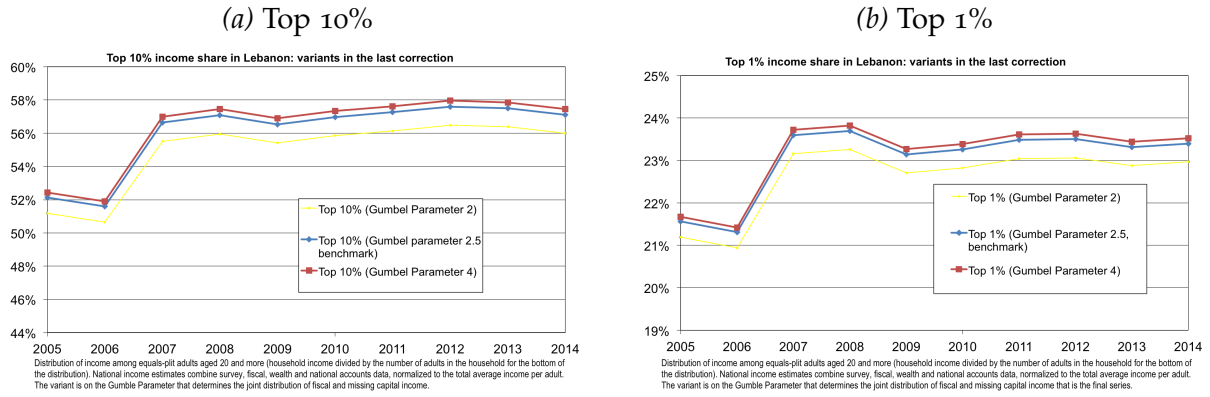


Figure A.7 – Variants for the Step 3: Effect of the billionaires' family size

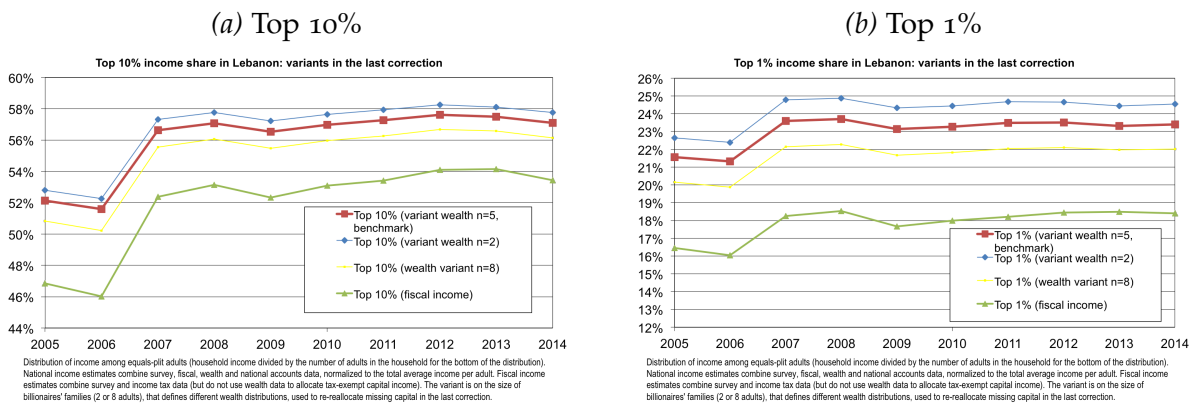


Figure A.8 – Main Tax form, Personal Income Tax

Republic of Lebanon Ministry of Finance Directorate General of Finance Directorate of Revenues – Income Tax		F1 (Individual) Page 1/4
Personal Declaration – Income Tax		
Full Name of taxpayer: (name) (father's name) (family name) Position: Please tick the appropriate case - <input type="checkbox"/> Lump-sum profit individual taxpayer - <input type="checkbox"/> Real profit individual taxpayer - <input type="checkbox"/> Partner taxpayer in a partnership Personal registration number (at the Ministry of Finance) Region of Activity: For the year: Is it the first declaration of the taxpayer? Yes No Is it the last declaration of the taxpayer? Yes No. If yes, please mention the reason:		
Marital status: <input type="checkbox"/> Single <input type="checkbox"/> Married <input type="checkbox"/> Divorced <input type="checkbox"/> Widow Number of dependent children:	Spouse: S/He works <input type="checkbox"/> Yes <input type="checkbox"/> No In case s/he works, personal registration number (at the Ministry of Finance):	
Personal Address Mohafazat Caza Region-Locality Real Estate Region District Street Building No of Lot/Section Floor Phone Fax Phone PO Box: Region Email:	Correspondence Address Mohafazat Caza Region-Locality Real Estate Region District Street Building No of Lot/Section Floor Phone Fax Phone PO Box: Region Email:	
The person who contributed to filling the declaration: Full name: Phone:		
Registration number (at the Ministry of Finance): Fax:		
Declaration contents:	Yes	No
* Profit (or Loss) from partnerships	<input type="checkbox"/>	<input type="checkbox"/>
* Profit (or Loss) from individual institutions and professions- Real profit	<input type="checkbox"/>	<input type="checkbox"/>
* Profit from individual institutions and professions – lump sum profit	<input type="checkbox"/>	<input type="checkbox"/>
* Estimated profit	<input type="checkbox"/>	<input type="checkbox"/>
* Salaries and wages revenues	<input type="checkbox"/>	<input type="checkbox"/>
* Statement of the institution/the profession's revenues and expenses F3 Form (obligatory for lump-sum profit)	<input type="checkbox"/>	<input type="checkbox"/>
* Statement of amounts paid to experts, accountants, consultants, lawyers, engineers, etc. (lump-sum profit taxpayers-F4 Form)	<input type="checkbox"/>	<input type="checkbox"/>
*Statement of deficit to be carried forward – for real profit taxpayers (F21)	<input type="checkbox"/>	<input type="checkbox"/>
*Statement of paid amounts to non-residents according to Articles 41, 42 and 43 for taxpayers on the basis of real profit and on the basis of lump-sum profit (F26)	<input type="checkbox"/>	<input type="checkbox"/>

(*) The natural or moral person is notified at the correspondence address given to the Tax Department. Therefore, read carefully Articles 27 and 28 of the tax procedures law No 44, dated 11/11/2008.

Figure A.9 – Tax form for labor income, Personal Income Tax

Republic of Lebanon Ministry of Finance Directorate General of Finance Directorate of Revenues – Tax on Wages and Salaries		R6 (Salaries)																																																																																
Individual Annual Statement of the Overall Revenues of the Employee/Wage-earner																																																																																		
Company/Institution name Commercial name Registration number (at the Ministry of Finance)	For the year Number of employees/wage-earners																																																																																	
Employee/Wage-earner's name Father's name Family name Personal registration number (at the Ministry of Finance) Type of work Type of wage* <input type="checkbox"/> Monthly <input type="checkbox"/> Daily <input type="checkbox"/> Hourly Family status* <input type="checkbox"/> Single <input type="checkbox"/> Married <input type="checkbox"/> Widow <input type="checkbox"/> Divorced Number of children: Number of people benefiting from family rebate** Work duration from (D/M/Y) ... / ... / ... to (D/M/Y) ... / ... / ... Number of working days for beneficiary from daily rebate:		Employee number ... out of ... (total number of employees)																																																																																
Employee/Wage-earner address: Mohafazat Caza Region/Locality District Street Building Floor Phone Phone PO Box Region Fax Email:																																																																																		
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 5%;">Description</th> <th style="width: 45%;">Total Revenues (1)</th> <th style="width: 20%;">Tax Exempted Revenues (2)</th> <th style="width: 30%;">Taxable Revenues (3)</th> </tr> </thead> <tbody> <tr><td>100</td><td>Main salary/daily wages</td><td></td><td></td></tr> <tr><td>110</td><td>Representation remuneration</td><td></td><td></td></tr> <tr><td>120</td><td>Bonuses, commissions and overtime</td><td></td><td></td></tr> <tr><td>130</td><td>Family compensation for the spouse</td><td></td><td></td></tr> <tr><td>140</td><td>Family compensation for the children</td><td></td><td></td></tr> <tr><td>150</td><td>Transportation compensation</td><td></td><td></td></tr> <tr><td>160</td><td>Car allowance</td><td></td><td></td></tr> <tr><td>170</td><td>Residence allowance</td><td></td><td></td></tr> <tr><td>180</td><td>Food allowance</td><td></td><td></td></tr> <tr><td>190</td><td>Clothing allowance</td><td></td><td></td></tr> <tr><td>200</td><td>Fund compensations</td><td></td><td></td></tr> <tr><td>210</td><td>Health insurances of all types</td><td></td><td></td></tr> <tr><td>220</td><td>Educational grants</td><td></td><td></td></tr> <tr><td>230</td><td>Marriage grants</td><td></td><td></td></tr> <tr><td>240</td><td>Birth grants</td><td></td><td></td></tr> <tr><td>250</td><td>Assistance in case of illness</td><td></td><td></td></tr> <tr><td>260</td><td>Assistance in case of death</td><td></td><td></td></tr> <tr><td>300</td><td>Other grants and benefits</td><td></td><td></td></tr> <tr><td>310</td><td>Total</td><td></td><td></td></tr> </tbody> </table>			Description	Total Revenues (1)	Tax Exempted Revenues (2)	Taxable Revenues (3)	100	Main salary/daily wages			110	Representation remuneration			120	Bonuses, commissions and overtime			130	Family compensation for the spouse			140	Family compensation for the children			150	Transportation compensation			160	Car allowance			170	Residence allowance			180	Food allowance			190	Clothing allowance			200	Fund compensations			210	Health insurances of all types			220	Educational grants			230	Marriage grants			240	Birth grants			250	Assistance in case of illness			260	Assistance in case of death			300	Other grants and benefits			310	Total		
Description	Total Revenues (1)	Tax Exempted Revenues (2)	Taxable Revenues (3)																																																																															
100	Main salary/daily wages																																																																																	
110	Representation remuneration																																																																																	
120	Bonuses, commissions and overtime																																																																																	
130	Family compensation for the spouse																																																																																	
140	Family compensation for the children																																																																																	
150	Transportation compensation																																																																																	
160	Car allowance																																																																																	
170	Residence allowance																																																																																	
180	Food allowance																																																																																	
190	Clothing allowance																																																																																	
200	Fund compensations																																																																																	
210	Health insurances of all types																																																																																	
220	Educational grants																																																																																	
230	Marriage grants																																																																																	
240	Birth grants																																																																																	
250	Assistance in case of illness																																																																																	
260	Assistance in case of death																																																																																	
300	Other grants and benefits																																																																																	
310	Total																																																																																	
Are rebated: 330 Family Rebate 340 Other Rebates																																																																																		
350 Net Revenues 360 Annual due tax																																																																																		

R6 Form – Tax of Wages and Salaries – 2010 Version
Website of the Ministry of Finance

and salaries – 2010 Version
Website of the Ministry of Finance

* Please tick the appropriate box. ** The figure includes the spouse in case she doesn't work and dependent children.

Figure A.10 – Tax form for self-employed individuals, liberal and independent professions

Republic of Lebanon Ministry of Finance Directorate General of Finance Directorate of Revenues – Income Tax		F3 (Individual)	
Statement of Revenues and Expenses of the Institution/the Profession For lump-sum profit taxpayers			
Name of the Taxpayer:	Institution/Profession's registration number (at the Ministry of Finance):		
Commercial Name:	Personal registration number (At the Ministry of Finance):		
For the year:			
<u>Address of the institution/the profession:</u> Mohafazat Caza Region-Locality District Street Real Estate Region No of Lot/Section Building Floor Phone Phone PO Box Region Fax Email			
Activity type	Revenues (*)	Profit rate	Net lump-sum profit
1	90	95	100
2	91	96	110
3	92	97	120
4	93	98	130
Total	94	99	140
Details of charges and expenses			
Sales of merchandise and raw and consumption material			150
Wages, salaries and other benefits			160
Employees and wage-earners insurance			170
Social security subscriptions			180
Commissions paid to third parties			190
Car and transportation expenses			200
Banking commissions, interests and expenses			210
Legal expenses, consultancies and similar expenses			220
Maintenance and repair expenses			230
Rent or investment			240
Other office expenses			250
Taxes, fees, and permits			260
Accommodation			270
Travelling expenses			280
Promotion and advertisement			290
Institution/Profession activity insurance expenses			300
Amortization**			310
Other expenses			320
Overall charges and expenses			330
Fixed assets buyouts			340
Amounts paid to non-residents			350

* If the taxpayer practices several activities within the same institution and did not separate the revenues of each activity, the highest rate applies to the lump-sum profit among these activities and on the overall revenues.

** For machinery, equipment, and furniture used in the institution in the first year, plus any addition and minus any wavering, the result is multiplied by the depreciation rate of each type of them.

Signature:

Figure A.11 – Tax form for other business incomes (1)

Republic of Lebanon Ministry of Finance Directorate General of Finance Directorate of Revenues – Income Tax				
Income Statement				F16 (Individual)
Line Number	Account description (1)	Current Financial Cycle		Previous Financial Cycle (4)
		Details (2)	Total (3)	
200	Goods sales			
210	Production sales			
220	Works sales			
230	Services sales			
240	Turnover			
250	Sold merchandise cost			
260	Sold production cost			
270	Works cost			
280	Services provision cost			
290	Overall cost			
300	Profit			
310	Consumer supplies cost			
External services				
320	Royalties			
330	Maintenance and repair			
340	Promotion and advertisement			
350	Transportation			
360	Rent			
370	Representation expenses			
380	Supervision expenses			
390	Travelling expenses			
400	Experts and consultants wages			
410	Insurance			
415	Paid commissions			
420	Other external services			
430	Total external services			
Employees wages/charges				
440	Salaries and wages			
450	Commissions			
460	Social security			
470	Transportation			
480	Employees meals			
490	Training expenses			
500	Insurances/Guarantees to users			
510	Other expenses			
520	Total employees wages/charges			
Corresponding taxes, fees and charges				
530	Corresponding taxes, fees and charges			
540	Fines			
550	Other administrative charges			
560	Total corresponding taxes, fees and charges			
Depreciation and investment provisions allocations				
570	Depreciation allocations			
580	Provisions allocations			
590	Total consumption and investment provisions allocations			
600	Total charges			

Figure A.12 – Tax form for other business incomes (2)

Republic of Lebanon Ministry of Finance Directorate General of Finance Directorate of Revenues – Income Tax				
Income Statement				F16 (Individual)
Line Number	Account description (1)	Current Financial Cycle		Previous Financial Cycle (4)
		Details (2)	Total (3)	
610	Recoveries from non-financial provisions			
620	Products with a fixed assets nature			
630	Other revenues			
640	Overall non-financial investment revenues			
650	Investment profit or loss			
660	Common operations dividends			
670	Placement and participation bonds revenues			
680	Net profit from placement bonds wavering			
690	Revenues from other movables			
700	Similar interests and revenues			
710	Positive exchange rate differences			
720	Recoveries from financial provisions			
730	Other financial revenues			
740	Overall financial revenues			
750	Institution's share out of the losses from joint operations			
760	Negative exchange rate differences			
770	Net charges on operation of placement bonds wavering			
780	Consumption and financial provisions allocations			
790	Financial institutions interests			
800	Financial institutions expenses			
810	Other interests			
820	Overall financial charges			
830	Net financial charges and revenues			
840	Losses resulting from wavering of fixed assets			
850	Grants and donations			
860	Other charges			
870	Overall other charges			
880	Profit resulting from wavering of fixed assets			
890	Other revenues			
900	Total other revenues			
910	Revenues of exceptional events			
920	Losses of exceptional events			
930	Differences due to changes in accounting policies			
940	Income before tax on profits			
950	Tax on profit			
960	Net income after tax (profit and loss)			

Table A.1 – From fiscal income to national income

	Type of income	Data sources
Labor Income	Mixed income (Self-employment income by independent)	Fiscal micro-files
	Non filers (including informal sector)	N.A
	Tax evasion	N.A
	Employer fringe benefits & payroll taxes	N.A
Capital Income	Other mixed income*	Fiscal micro-files
	Corporated profits	Government tax reports
	<i>incl. Undistributed profits (retained earnings)</i>	
	<i>incl. Distributed profits</i>	Government tax reports
	Interest incomes	National accounts
	Rental Income	Fiscal micro-files
	<i>incl. built property revenues</i>	
	<i>incl imputed rents and property tax</i>	Government tax reports
<i>incl. royalties</i>	N.A	
Non filers and others	N.A	

Notes: Statistics on the distribution of income expressed in PPP Euro 2016. Adult individual aged 20 and more; Equal-split assumption among adult members of a household. In 2016, 1 euro = 1641 LBP (market exchange rate) or 172.7 pound (PPP). Income corresponds to pre-tax national income. Fractiles are defined relative to the total number of adult individuals in the population. Corrected estimates (combining survey, fiscal, wealth and national accounts data), from 2014 adjusted for the price change between 2014-2016 (shares are not affected).

*Made by partners in partnerships and individuals in S-corporations.

B. APPENDIX TO CHAPTER 2

B.1 Survey data - sources

In order to estimate the entire income distribution of the Middle East, we first gather available income data to estimate the raw survey distribution for each country, over the period. Table A1 summarizes the survey sources used in this paper. When we could not access micro-data, we use survey income tabulations extracted from survey reports published online by the national statistical offices. Raw statistical data are given the directory "HouseholdSurveyData". The Stata format do-files used to generate the raw survey distributions can be found in Gpinter/Do/IncomeSurveySeries.

Egypt, Iraq, Jordan and Palestine

We have access to survey micro-data for seven countries. For Egypt, Iraq, Jordan, and Palestine we use the harmonized databases created by the Economic Research Forum (OAMDI initiative) and take the households total disposable income as income concept. For some years and countries, we have additional variables (as the expenditure on taxes and levies, or imputed rental income) that we include to approach the pre-tax, post-replacement income concept defined in Alvaredo et al. (2016). All computations can be found in each country specific do file, in Gpinter/Do/IncomeSurveySeries. The variables available are presented in the sheet "ERFHarmonization", in the excel file "CleaningMicroData.xlsx".

Turkey

For Turkey, we use the Household Budget Survey micro-data for the years 1994, 2002 and 2013. As explained in the main paper, the Turkish micro-data are more detailed so that we can adjust the income variable to approach the pre-tax, post-replacement income concept. The sheet "TurkeyHBSIndivHarmonization", in "CleaningMicroData" presents how we computed the income variable. For the years 2014-2016, we use tabulations on per capita income computed with the Income and Living Conditions Survey and available online on the Turkish Statistical Institute website. The raw data and computations can be found in the directory "HouseholdSurveyData/Turkey".

Iran

For Iran, we used micro-data publicly available online on the website of the Statistical Center of Iran, for the years 2010 and 2013. The micro data distinguish between Rural and Urban areas. We therefore used the Gpinter "merge option" to infer the national income distribution per adult. The raw data and computations can be found in the directory "HouseholdSurveyData/Iran".

Gulf countries

For Gulf countries, only tabulations published by national statistical offices are available. They systematically distinguish between "National" and "Foreigner" populations.

Using the Gpinter interface, we derive the per adult income distribution for each sub-group of the population over the period. The merge option enables us to correct for sampling issues (noted in the Kuwaiti data by El Katiri et al., 2011, but that we find in other Gulf countries) and the fact that foreigners are systematically under sampled in survey data. We directly weight the distribution by the adult population of each sub-group. Details about the estimation of the share of foreigners in the total population are presented in the excel file "GulfCountries.xlsx" (directory HouseholdSurveyData). Table A2 displays inequality statistics derived from survey data for foreigners, nationals and the entire population of Gulf countries. Taking into account foreigners increases inequality in all countries and for all years. Note however that the distinction between the two populations can only be done at the survey distribution level, as we afterwards normalize the series to the national average income per adult and lose track of the differential in average income between the two populations. We therefore still highly underestimate income inequality within these countries.

Saudi Arabia

Three household surveys have been conducted in Saudi Arabia, in 1999, 2007 and 2013, but the reports published by the statistical office do not contain exploitable information on the income or expenditure distribution. We therefore proceed as follow. First, we take the average standardized distributions of income computed with available survey data in other Gulf countries, that is we divide all thresholds and bracket averages for all 127 generalized percentiles by average wealth, and we compute the arithmetic average for the five countries, over the period. It seems unjustified to adjust this average Bahrain-Kuwait-Oman-Qatar-UAE normalized distribution to Saudi Arabia directly as it would imply that Saudi Arabia inequality levels follow the same trends as the other countries. Given its weight in the Middle East in terms of population and income, we attribute the most equal Bahrain-Kuwait-Oman-Qatar-UAE normalized distribution (the year 2008) that we adjust to the Saudi per adult national income over the period. Note however that this procedure is rather conservative, given that published Gini coefficients (51.3 and 45.9 in 2007 and 2013 for the Saudi population) are greater than what is observed in other countries (see Table A2).

Countries with information on expenditure only: Syria, Yemen, and Kuwait

Finally, for Syria Yemen and Kuwait, only information on expenditure is available. We therefore proceed as follow. First, we use the Palestinian micro-data, for which both household income and expenditure are available, to compute the residual saving rates by g-percentile, that is, for each g-percentile, we compute the ratio of income over expenditure and we apply these ratio to the Syrian and Yemeni expenditure distributions. For Kuwait, we first derive the ratios between income and expenditure by g-percentile using Bahraini and Emirati tabulations, for which tabulations are available for both the income and expenditure distributions. We compute these correction coefficients for the foreigner and national populations and apply them to the expenditure distribution of each sub-group of the Kuwaiti population.

Other issues: unit of observation and cleaning procedure

Two additional remarks on the procedure are in order. First, as explained in the main paper, we take as unit of observation the adult individual, aged 20 years and more. For all countries with micro-data, with the exception of Iran, we estimate the number of adults in each household, using information on the household type or the age of household members in the case of Turkey. We then normalize the total adult population in the survey data to the demographic figures gathered in WID.world and deduce the income per adult. For Iran, we divide household income by the household size multiplied by an estimate of the adult population in each area (Urban and Rural). Finally, for countries with tabulations only, the number of adults by income bracket is rarely available. We therefore divide household income in each group by the average household size multiplied by the share of adults in the entire economy, assuming that the income distribution per adult is the same as the household distribution. We do the same when the average household size per bracket is available. The second issue concerns the cleaning of the survey micro-data. All hypotheses made and comparison between variants can be found in “CleaningMicroData.xlsx”, in the directory “HouseholdSurveyData”. For each country, we dismiss the observations for which the ratio between the income and the average income is greater than what is observed on average for other years and countries.

B.2 Income and wealth distribution series

B.2.1 Income distribution series

The general methodology to construct our income distribution series is summarized in the main paper (Section 3) and consists of three steps: in step 1 we use survey data and generalized Pareto interpolation techniques (Blanchet et al., 2022) to estimate survey income distributions for each country; in step 2 we use the Lebanese fiscal data to correct the top of these distributions and obtain corrected estimates of the distribution of fiscal income, by g-percentile; in step 3 we use national accounts and wealth data in order to include tax-exempt capital income (such as undistributed profits, imputed rent and other “non-fiscal income”) and to obtain corrected estimates of the distribution of pre-tax national income by g-percentile. We then aggregate each country specific distribution using the merge option in the Gpinter interface.

In the following, we discuss a number of additional issues about variant series and robustness checks. We focus on the regional distribution. The effect of each correction on within-country inequalities is broadly the same as in Lebanon, to a large extent by construction and details about variants on national distributions can be found in Assouad (2021).

Figures B.1-B.4 display the main results of the papers and the impact of the two corrections. Two facts stand out. First, the impact of the fiscal correction is larger than the one of the wealth correction. Second, both corrections do not change the trends. Furthermore, as one can see by comparing Tables A1, ?? and ??, the difference in magnitude of the fiscal correction for each country comes, first, from the initial survey inequality levels and second, from the ratio between total survey income and total national income, which

can be very low by international standards for some countries (Table A1). Given that, we focus on variants on the fiscal correction when we aggregate the series.

Fiscal Data Correction

Regarding the fiscal-data correction, we choose in our benchmark series to apply the same average upgrade factors by g-percentile (estimated using the Lebanese fiscal data over the 2005-2014 period) to the entire 1990-2016 period. In the absence of alternative data to correct the top of the survey distributions, it seems reasonable to apply the Lebanese correction coefficients to other countries. Nevertheless, it would clearly be unjustified to assume that the correction follows the same trends as Lebanon.

Details about the Lebanese fiscal micro-files, their limits and how to combine them with survey data are provided in Assouad (2021). In particular, the fiscal micro files are based upon a “taxable income” concept (i.e. income subject to income tax, after a number of deductions allowed by tax legislation) that may be significantly smaller than “fiscal income” (defined as the sum of all income items legally subject to taxation, before any deduction) due to large deductions and exemptions. We therefore compute the fiscal correction coefficients by assuming that the ratio between taxable income and fiscal income is equal to $r=80\%$. Figures B.5-B.8 display variants with $r = 60\%$ and $r = 90\%$ for the Middle East series and for Lebanon. The impact at the country level is greater than at the regional level, or in other words, the effect of the correction and the hypotheses made does not cumulate when we aggregate countries together.

Figures B.9-B.10 show the effect of the correction profile chosen to compute the correction coefficients (that is to link the survey data at the bottom to the fiscal data at the top of the distribution). The profile chosen does not have a large impact on income shares.

The way we estimate our upgrade factors is described in files CompCorrectionCoeffLeb.xlsx. We obtain the file CorrectionCoefficientsAllCountries.xlsx that is used to apply the correction to all countries. Figures B.11 and B.12 display the Pareto curves of the final series computed with different profiles, and figures B13 the Pareto curves of our benchmark series for selected years. Our correction leads to smooth curves comparable to what is observed in other countries.

Missing capital income correction

As explained in the main paper (section 2.2), we proceed to a second correction accounting for the missing capital incomes. There are two steps. The first one is to estimate the amount of capital income missing. For Lebanon, they are estimated using national accounts and public finance reports and represent 20% of national income. For other countries, in absence of better data, we assume that they represent 10% of national Income, a rather conservative amount given the low ratios between total fiscal income and total national income observed (see excel file ShareNationalIncome.xlsx). The second step consists in re-allocating these amounts to estimate the distribution of total personal income. To do so, we first assume that they follow the same distribution as wealth, that we estimate using billionaires data. Second, we use the family of Gumbel copulas to estimate the joint distribution of fiscal income and missing capital income.

The effect of each hypothesis made during the last step on national distributions is explained in detail for the Lebanese case in [Assouad \(2021\)](#). We nevertheless provide robustness check on hypothesis on the size of the billionaires' family to estimate the wealth distribution that is used to reallocate the missing capital income (see section B.2 below). As shown in Figures [B.14](#) and [B.15](#), this has a small impact on the series, smaller than the effect on each national distribution.

Finally, Figures [B.16](#) and [B.17](#) show the series according to alternative definition for the Middle East (without Turkey, without Iran or without Gulf countries). For the two first, this has no effect on our main conclusion. As the main driving force behind the extreme concentration of income in the Middle East is the differential in per adult income between oil-rich countries and other population-rich countries, our series without Gulf countries are way below our benchmark estimates. Note that the level of inequality stays nevertheless very high with the top 1% and top 10% of the population accounting for almost 25 and 56% of total income respectively.

B.2.2 Wealth distribution series

Regarding the distribution of missing capital income, we assume it follows the same distribution as wealth. Wealth data are very limited in the Middle East. Unlike other countries, where we can use a combination of sources and methods, all we have in the region at this stage is billionaire data. The estimation procedure is the same for all countries and described in detail in [Assouad \(2021\)](#). We discuss here some issues concerning rich list data.

To construct our billionaires' data, we combine two main sources: Forbes and Arabian Business Rich Lists. With the exception of Lebanon, billionaires' wealth does not seem particularly high in the Middle East by international standards. Rich Lists nevertheless do not take into account the wealth of state rulers, which may create a bias particularly significant in the region. We include additional sources on leaders' wealth (see excel file "BillionairesData.xlsx, sheets "ForbesRoyalsDictators" and "OtherSources"). The data remain however limited, as state leaders' wealth included do not cover all ruling families and are often available for limited time periods (for example, we have information on the Assad family's wealth for two years only). This leads to very variable estimates of the share of billionaires' wealth to national income, which do not display any trends. We therefore use the average share of billionaires' wealth in national income over the period to estimate the 2016 wealth distributions, that we adjust to the average wealth per adult in the period 1990-2016 period (see excel file BillionairesData.xlsx, sheet "Stata"). For countries without billionaires' data -Iran Jordan Palestine Yemen - we simply upgrade the average standardized US-France-China distribution of wealth to each country specific average wealth, estimated as described in [Assouad \(2021\)](#) by multiplying the national income by the average wealth/income ratio for countries with available data reported in WID.world.

B.3 Additional Figures and Tables

Figure B.1 – Decomposing the level of Middle East top 10% income share

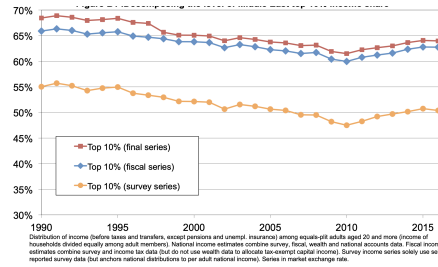


Figure B.2 – Decomposing the level of Middle East top 1% income share

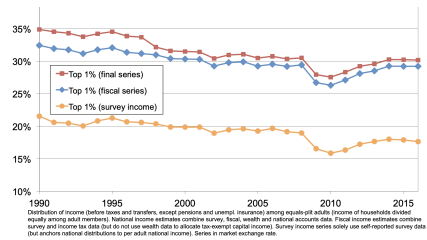


Figure B.3 – Decomposing the level of Middle East Middle 40% income share

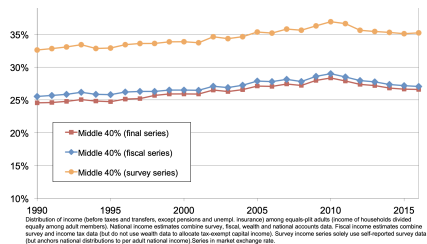


Figure B.4 – Decomposing the level of Middle East Bottom 50% income share

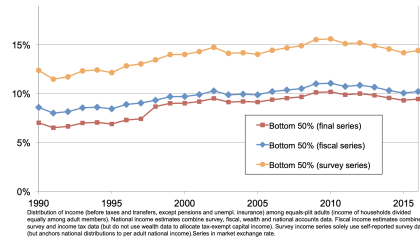


Figure B.5 – Top 10% income share in the Middle East, 1990-2016: impact of the tax correction (1)

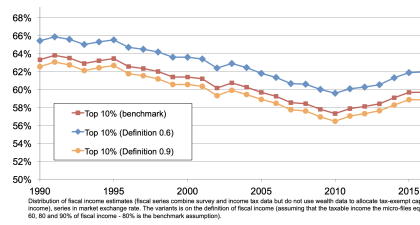


Figure B.6 – Top 1% income share in the Middle East, 1990-2016: impact of the tax correction (1)

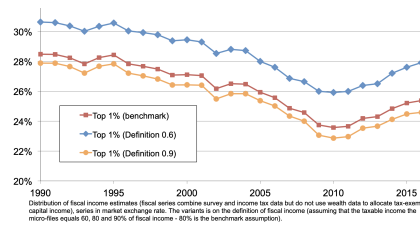


Figure B.7 – Top 10% income share in the Middle East, 1990-2016: impact of the tax correction (1)

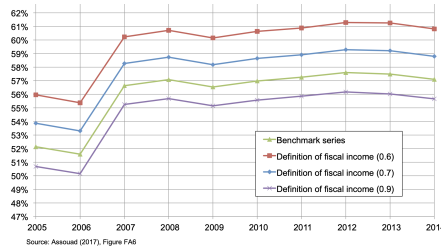


Figure B.8 – Top 1% income share in the Middle East, 1990-2016: impact of the tax correction (1)

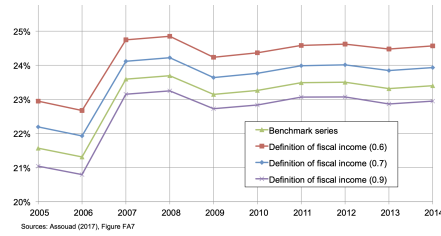


Figure B.9 – Top 10% income share in the Middle East, 1990-2016: impact of the tax correction (2)

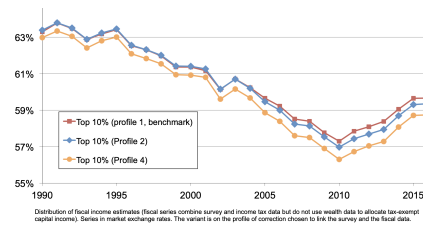


Figure B.10 – Top 1% income share in the Middle East, 1990-2016: impact of the tax correction (2)

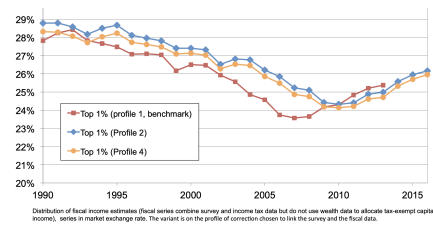


Figure B.11 – Pareto coefficients curves in the Middle East, 1990: impact of tax correction (2)

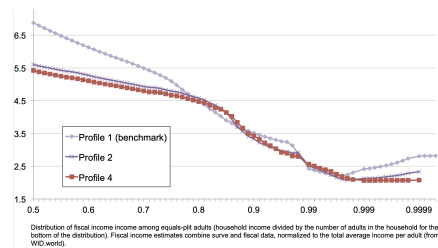


Figure B.12 – Pareto coefficients curves in the Middle East, 2016: impact of tax correction (2)

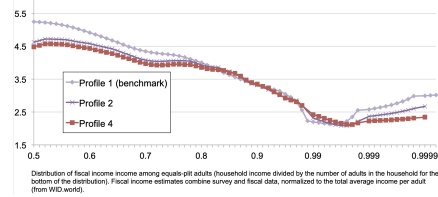


Figure B.13 – Inverted Pareto curves in the Middle East in selected years, final benchmark distribution

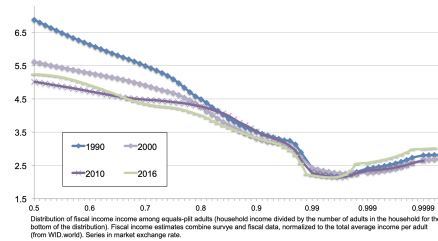


Figure B.14 – Top 10% income share in the Middle East, 1990-2016: impact of the wealth correction

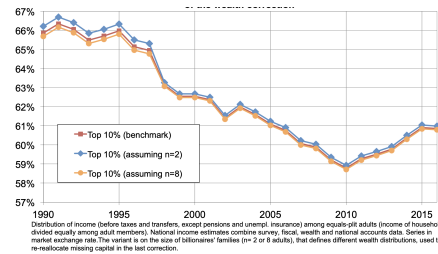


Figure B.15 – Top 1% income share in the Middle East, 1990-2016: impact of the wealth correction

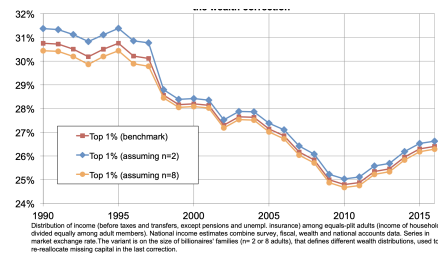


Figure B.16 – Top 10% income share in the Middle East, 1990-2016, geographical variants

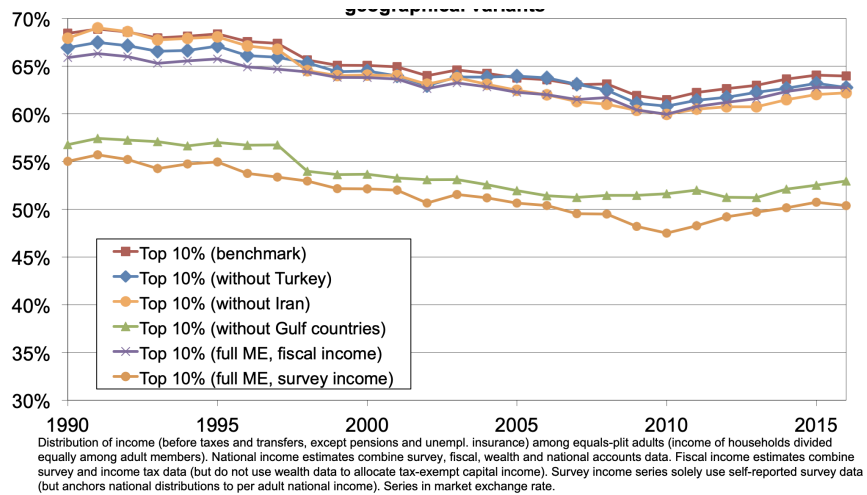


Figure B.17 – Top 1% income share in the Middle East, 1990-2016, geographical variants

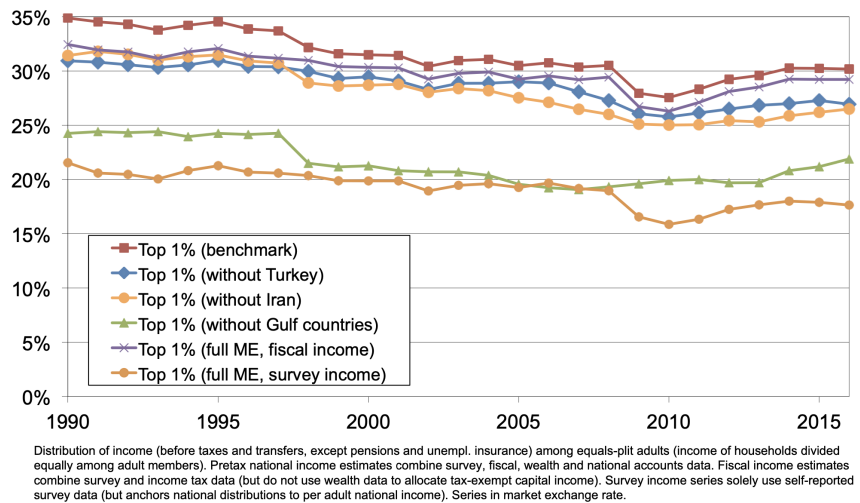


Table A1 – Household surveys used in this paper (1990-2016)

		Survey years	Average ratio (total survey income)/(total national income)	Format
Turkey		1994, 2002-2016	43 %	micro-data and tabulation
Iran		2010, 2013	49 %	micro-data
Egypt		1999, 2004, 2008, 2010, 2012, 2015	40 %	micro-data
Iraq-Syria-Other (non-Gulf)		1992-2013	53 %	
Iraq	2012	2007	60 %	micro-data
Syria	2012	2004	56 %	micro-data
Jordan	2012	1992, 2002, 2006, 2008, 2010, 2013	70 %	micro-data and tabulation
Lebanon	2012	2007	37 %	tabulation
Palestine		1996-1998, 2004-2008, 2010-2011	65 %	micro-data
Yemen	2012	2006	33 %	micro-data
Gulf Countries		1995-2013	30 %	
Saudi Arabia	2012	N/A	30 %	N/A
Oman	2012	2010	29 %	tabulation
Bahrain	2012	1995, 2005, 2015	37 %	tabulation
UAE	2012	1998, 2009	39 %	tabulation
Kuwait	2012	2007, 2013	21 %	tabulation
Qatar	2012	2007, 2012	23 %	tabulation

Table A2 – Gulf-Countries Income Inequality Series: nationals vs foreigners in the survey data

Country	Year	Group	Bottom 50%	Middle 40%	Top 10%	Top 1%	Gini	P10/ average	P50/ average	P90/ average	P99/ average	Pareto b(10%)	Pareto b(50%)	Pareto b(90%)
Bahrain	1995	Nationals	24 %	45 %	30 %	7 %	39 %	34 %	75 %	168 %	487 %	3,2	2,0	1,7
		Foreigners	21 %	41 %	38 %	8 %	46 %	28 %	70 %	197 %	535 %	3,9	2,3	1,8
		Total	23 %	43 %	34 %	7 %	42 %	29 %	72 %	179 %	511 %	3,7	2,2	1,8
	2005	Nationals	30 %	44 %	26 %	6 %	31 %	47 %	88 %	159 %	398 %	2,3	1,6	1,6
		Foreigners	20 %	44 %	37 %	7 %	46 %	34 %	60 %	197 %	585 %	3,2	2,7	1,8
		Total	23 %	45 %	32 %	7 %	40 %	33 %	71 %	178 %	503 %	3,3	2,2	1,7
	2015	Nationals	30 %	44 %	26 %	6 %	31 %	47 %	88 %	159 %	398 %	2,3	1,6	1,6
		Foreigners	20 %	44 %	37 %	7 %	46 %	34 %	60 %	197 %	585 %	3,2	2,7	1,8
		Total	23 %	45 %	32 %	7 %	41 %	33 %	71 %	180 %	515 %	3,3	2,2	1,7
Kuwait	2007	Nationals	31 %	47 %	23 %	4 %	28 %	51 %	83 %	158 %	318 %	2,1	1,7	1,4
		Foreigners	31 %	44 %	25 %	6 %	29 %	50 %	88 %	154 %	387 %	2,1	1,6	1,6
		Total	22 %	44 %	34 %	7 %	43 %	36 %	63 %	202 %	487 %	3,0	2,5	1,6
	2013	Nationals	34 %	45 %	21 %	5 %	24 %	58 %	89 %	145 %	293 %	1,8	1,5	1,4
		Foreigners	36 %	41 %	23 %	5 %	23 %	54 %	94 %	134 %	349 %	2,0	1,4	1,6
		Total	22 %	43 %	35 %	7 %	44 %	33 %	56 %	227 %	468 %	3,3	2,8	1,5
Oman	2010	Nationals	20 %	49 %	31 %	6 %	43 %	26 %	77 %	199 %	443 %	4,2	2,1	1,5
		Foreigners	13 %	44 %	42 %	10 %	56 %	16 %	55 %	219 %	677 %	7,0	3,1	1,9
		Total	17 %	48 %	35 %	7 %	50 %	12 %	64 %	213 %	519 %	9,3	2,6	1,6
Qatar	2007	Nationals	28 %	47 %	25 %	5 %	33 %	34 %	79 %	172 %	353 %	3,2	1,8	1,4
		Foreigners	22 %	50 %	28 %	5 %	41 %	18 %	81 %	196 %	374 %	6,1	1,9	1,4
		Total	20 %	48 %	31 %	6 %	45 %	25 %	71 %	198 %	460 %	4,4	2,2	1,5
	2012	Nationals	28 %	47 %	24 %	4 %	32 %	40 %	86 %	170 %	327 %	2,7	1,7	1,4
		Foreigners	27 %	47 %	26 %	4 %	34 %	36 %	82 %	175 %	338 %	3,0	1,8	1,4
		Total	23 %	44 %	33 %	7 %	41 %	31 %	73 %	181 %	522 %	3,5	2,1	1,7
UAE	1998	Nationals	27 %	47 %	26 %	5 %	34 %	37 %	86 %	164 %	376 %	2,9	1,7	1,5
		Foreigners	20 %	43 %	37 %	8 %	46 %	28 %	64 %	199 %	554 %	3,9	2,5	1,8
		Total	19 %	46 %	35 %	7 %	46 %	26 %	63 %	209 %	521 %	4,2	2,5	1,6
	2009	Nationals	20 %	52 %	27 %	4 %	42 %	28 %	68 %	215 %	332 %	3,9	2,3	1,3
		Foreigners	18 %	45 %	37 %	5 %	48 %	34 %	55 %	234 %	493 %	3,2	3,0	1,5
		Total	17 %	45 %	38 %	6 %	50 %	30 %	53 %	233 %	497 %	3,6	3,1	1,6

Table A3 – Country-Level Income Inequality Series: Income Shares and Other Indicators for the Survey Distribution

Country	Year	Bottom 50%	Middle 40%	Top 10%	Top 1%	Gini	P10/ average	P50/ average	P90/ average	P99/ average	Pareto b(10%)	Pareto b(50%)	Pareto b(90%)
Bahrain	1995	23 %	43 %	34 %	7 %	42 %	29 %	72 %	179 %	511 %	3,7	2,2	1,8
	2005	23 %	45 %	32 %	7 %	40 %	33 %	71 %	178 %	503 %	3,3	2,2	1,7
	2015	23 %	45 %	32 %	7 %	41 %	33 %	71 %	180 %	515 %	3,3	2,2	1,7
Egypt	1999	26 %	43 %	30 %	8 %	36 %	41 %	76 %	163 %	481 %	2,6	1,9	1,8
	2004	27 %	44 %	28 %	7 %	35 %	41 %	78 %	164 %	448 %	2,6	1,9	1,7
	2008	28 %	44 %	27 %	7 %	33 %	44 %	80 %	162 %	425 %	2,4	1,8	1,6
	2010	29 %	45 %	26 %	6 %	31 %	45 %	83 %	161 %	400 %	2,4	1,7	1,5
	2012	29 %	45 %	25 %	6 %	31 %	45 %	83 %	160 %	373 %	2,4	1,7	1,5
	2015	29 %	44 %	27 %	7 %	33 %	44 %	81 %	157 %	432 %	2,4	1,8	1,7
Iran	2010	22 %	47 %	31 %	7 %	41 %	27 %	76 %	183 %	458 %	4,0	2,0	1,6
	2013	25 %	46 %	29 %	6 %	37 %	34 %	79 %	176 %	431 %	3,2	1,9	1,6
Iraq	2007	24 %	44 %	32 %	10 %	40 %	32 %	76 %	168 %	511 %	3,3	2,0	1,8
Jordan	1992	24 %	45 %	31 %	9 %	40 %	32 %	75 %	172 %	478 %	3,4	2,0	1,7
	2002	26 %	47 %	27 %	6 %	35 %	37 %	80 %	174 %	378 %	2,9	1,9	1,5
	2006	26 %	47 %	28 %	6 %	36 %	35 %	81 %	180 %	392 %	3,1	1,8	1,5
	2008	27 %	46 %	28 %	6 %	35 %	39 %	81 %	172 %	396 %	2,8	1,8	1,6
	2010	26 %	43 %	31 %	9 %	37 %	36 %	78 %	158 %	498 %	3,0	1,9	1,9
	2013	26 %	46 %	28 %	6 %	36 %	35 %	81 %	175 %	401 %	3,1	1,8	1,5
Kuwait	2007	22 %	44 %	34 %	7 %	43 %	36 %	63 %	202 %	487 %	3,0	2,5	1,6
	2013	22 %	43 %	35 %	7 %	44 %	32 %	55 %	226 %	467 %	3,3	2,8	1,5
Lebanon	2005	20 %	46 %	33 %	8 %	45 %	27 %	72 %	196 %	486 %	4,1	2,2	1,6
	2007	20 %	47 %	33 %	8 %	46 %	23 %	70 %	203 %	499 %	4,7	2,3	1,6
Oman	2010	17 %	48 %	35 %	7 %	50 %	12 %	64 %	213 %	519 %	9,3	2,6	1,6
Palestine	1996	22 %	49 %	29 %	5 %	41 %	26 %	77 %	197 %	400 %	4,2	2,0	1,4
	1997	22 %	49 %	29 %	5 %	41 %	26 %	77 %	197 %	400 %	4,2	2,0	1,4
	1998	22 %	49 %	29 %	5 %	41 %	26 %	77 %	197 %	400 %	4,2	2,0	1,4
	2004	21 %	48 %	30 %	6 %	42 %	24 %	77 %	188 %	439 %	4,5	2,0	1,5
	2005	20 %	49 %	31 %	6 %	44 %	23 %	74 %	204 %	438 %	4,8	2,1	1,5
	2006	21 %	48 %	31 %	6 %	43 %	24 %	74 %	192 %	444 %	4,6	2,2	1,6
	2007	19 %	48 %	33 %	7 %	46 %	21 %	70 %	203 %	479 %	5,2	2,3	1,6
	2008	20 %	49 %	31 %	6 %	44 %	24 %	72 %	198 %	434 %	4,6	2,2	1,5
	2010	20 %	48 %	33 %	7 %	45 %	22 %	71 %	200 %	475 %	4,9	2,3	1,6
	2011	20 %	48 %	31 %	6 %	44 %	23 %	73 %	199 %	439 %	4,7	2,2	1,5
	Qatar	2007	20 %	48 %	31 %	6 %	45 %	25 %	71 %	198 %	460 %	4,4	2,2
2012		23 %	44 %	33 %	7 %	41 %	31 %	73 %	181 %	522 %	3,5	2,1	1,7
SaudiArabia	2008	20 %	46 %	34 %	7 %	45 %	27 %	65 %	205 %	490 %	4,1	2,5	1,6
Syria	2004	21 %	49 %	30 %	5 %	42 %	25 %	75 %	200 %	343 %	4,3	2,1	1,4
Turkey	1994	21 %	44 %	34 %	9 %	44 %	28 %	70 %	180 %	559 %	3,9	2,2	1,8
	2002	22 %	43 %	35 %	10 %	43 %	29 %	70 %	177 %	584 %	3,7	2,2	1,9
	2003	23 %	44 %	34 %	9 %	42 %	30 %	71 %	174 %	576 %	3,6	2,2	1,9
	2004	23 %	45 %	32 %	9 %	41 %	29 %	74 %	177 %	496 %	3,7	2,1	1,7
	2005	24 %	46 %	30 %	7 %	40 %	30 %	78 %	177 %	467 %	3,6	2,0	1,6
	2006	25 %	46 %	29 %	7 %	38 %	33 %	78 %	174 %	458 %	3,3	1,9	1,6
	2007	25 %	46 %	29 %	6 %	37 %	33 %	80 %	173 %	425 %	3,3	1,9	1,6
	2008	25 %	46 %	30 %	7 %	38 %	33 %	79 %	171 %	434 %	3,3	1,9	1,7
	2009	24 %	45 %	31 %	8 %	40 %	31 %	76 %	172 %	523 %	3,5	2,0	1,7
	2010	25 %	45 %	30 %	8 %	39 %	33 %	77 %	172 %	472 %	3,3	2,0	1,7
	2011	24 %	45 %	31 %	8 %	39 %	33 %	76 %	174 %	486 %	3,3	2,0	1,7
	2012	25 %	45 %	31 %	8 %	39 %	34 %	76 %	173 %	459 %	3,2	2,0	1,7
	2013	25 %	44 %	30 %	7 %	38 %	36 %	77 %	168 %	446 %	3,0	2,0	1,7
	2014	24 %	45 %	31 %	9 %	40 %	34 %	77 %	168 %	478 %	3,3	2,0	1,8
	2015	23 %	45 %	32 %	9 %	40 %	33 %	76 %	168 %	489 %	3,3	2,0	1,8
	2016	23 %	44 %	33 %	10 %	41 %	33 %	75 %	164 %	507 %	3,3	2,1	1,9
UAE	1998	19 %	46 %	35 %	7 %	46 %	26 %	63 %	209 %	521 %	4,2	2,5	1,6
	2009	17 %	45 %	38 %	6 %	50 %	30 %	53 %	233 %	497 %	3,6	3,1	1,6
Yemen	2006	21 %	49 %	30 %	5 %	42 %	25 %	76 %	203 %	349 %	4,4	2,1	1,4

C. APPENDIX TO CHAPTER 3

C.1 Data sources

Atatürk Visits

The main source used to build my treatment variable is a secondary source, the book by [Kocatürk \(1988\)](#). I complement this source with academic articles—mostly in Turkish—analyzing the visits in a given city and region, listed in the Reference section of the Appendix and municipalities’ websites, listed in Figure C.1. Finally, I cross-check the information by looking at historical newspapers, available online on the website of the project “[A look at History using newspapers](#)”, implemented by Istanbul University. Figure C.2 provides example of historical articles describing the visits. Using these various sources, I collect information on the date and location of the universe of Atatürk’s visits, listed in Table A2, and can document the activities he conducted for 122 out of 154 visits. The following section provides several examples of visits to illustrate the activities implemented.

- **Tarsus, March, 1923.** “The second night he spent in Tarsus, Mustafa Kemal asked to meet with the hosts of the house he was staying in with his delegation. Their host were Doctor Ali Refik and his wife Nimet Hanim. Kemal asked whether they had children. Ali Refik brought him their two daughters. Kemal asked the girls their names and when he heard they were named Güzin and Umran, he said “*These are Arabic names, we should no longer use them. Instead, we will use Turkish names. Shall the name of Güzin become Gazne, and of Umran, Turan*”. Kemal also explained them why it was necessary to purify the language”.¹

- **Kastamonu and Inebolu August, 1925** “On August, 23rd, Atatürk left Ankara for Kastamonu, with Nuri (Conker), deputy for Konya, Fuat (Bulca), deputy for Rize, his secretary general and other members of his government. He was welcomed by twenty one gun fires. The next day, he visited military bases, wearing his uniform. Afterwards, he inspected the hospital and the public library. Then, he went to the municipality building where he met several delegations of Turkish Hearths from other nearby districts. The evening, the inhabitants of Kastamonu organized a torchlight procession in his honor. Atatürk and his delegation then went to Inebolu on the 25th of August, 1925. Atatürk met with local government officers, artisans, military officers and representatives of the Turkish Hearth during two days. On August, the 27th of 1925, he gave a speech in Inebolu’s Turkish Hearth hall, which came to be known as the “Hat Speech”. He promoted hats along with trousers and shirts as the “civilized” and “westernized” way of dressing while arguing that wearing a fez, the traditional Ottoman hat, was backward and not worth of a Turk”.²

- **Sivas, September, 1928:** “During his stay in Sivas, Mustafa Kemal held a meeting in the city’s public square, to speak about the new alphabet and check the progress on its use. He invited a butcher from the crowd, Abidin, to the blackboard. Abidin

1. Source available [online](#).

2. Source available [online](#).

came to the front and admitted that he did not know the new alphabet. In ten minutes, Kemal taught him the vowels and showed him how to write his name. As Abidin learnt a new letter, repeated rounds of applause were shaking the square. When Abidin finally was able to write his name, the square erupted in cheers and shouted: “*Long live Turkey and the great Atatürk !*”. Figure 3.2 (a) is a photography of the corresponding scene, where Kemal describes the new latin alphabet to the crowd.³

- **Aydın, February, 1931** “Following the Menemen Incident, a revolt led by Dervish Mehmet Efendi to protest against the secularizing reforms and calling for the restoration of the Sharia Law and the Caliphate, Atatürk visited the Aegean region. On February the 3rd, 1931 at 4:30pm, Atatürk arrived in Aydın. He visited the municipality building and the Turkish Hearth.⁴ He gave a speech to the members of the Hearth, in where he stressed the importance of their role in promoting and teaching his reforms: “(...)The Turkish Hearths are the cultural branches of the Republican’s People’s Party. The party will educate the people in all fields, such as science, economics, politics and fine arts. Members of the Turkish Hearths should explain the Republican People’s Party program to the people.”⁵

- **Gaziantep, January, 1933:** “On January, the 25th, 1933, Atatürk went to Gaziantep. After a long journey, he first stoped in Fevzipaşa, a village in the West of Gaziantep. He was greeted by Gaziantep’s Governor, Major, and the party’s provincial representative. He arrived to the city center the next day, on the day of Eid-el-Fitr. The crowd celebrated his arrival with a flourish of trumpets. He went to the governor’s office to celebrate the Eid. Afterwards, he went to the municipality, where he met with local elites and listened to their concerns. When he was informed about the need for a high school, he immediately took action and a part of the Gaziantep secondary school was turned into a high school within five days”.⁶

- **Yozgat, February, 1934:** “Atatürk spent the night of 2 February 1934 in his train at the train station of Yerköy, a district of Yozgat province. The next day he arrived in Yozgat city center at 16:30. During his presence in Yozgat, he visited the government’s house, military bases, his party’s office, People’s House, the municipality, the high school and governor’s residence respectively. He finished all his visits in nine hours to leave the city at 1:30 AM and passed the night at his train in Yerköy.

- **Elaziz, November, 1937** “As part of his Eastern campaign tour, Atatürk went to Elaziz. On his way, he made a stop in Sivrice, a town next to the lake Gölçük,

3. Source: Yildirim (1996).

4. Source: Newspaper Milliyet, 5.2.1931, 1.

5. Source: Newspaper “Vakit”, 5.2.1931 and Günver Güneş “Mustafa Kemal Atatürk’ün Aydın Seyahatleri”, Atatürk Aracştırma Merkezi Dergisi 21 (2020).

6. Source: Newspaper Milliyet, 27.1.1933, 1.

to rest. He was particularly happy to hear that the surrounding mountains were named after the Khazars, a Turkic State of Central Asia. Kemal seized the occasion to rename the lake Gölcük to Khazar. On November, the 17th, 1937, Atatürk and his delegation arrived to the Elaziz district. He visited the People's House at Pertek. On his way to Pertek, Atatürk inaugurated a bridge which he renamed "Singeç", instead of as "Soyungeç" or Sungeç" arguing that this name was the most compatible with the Turkish pronunciation. On the same night, a celebration was organized in his honor at Elaziz's People's House. Two speeches were given, the first by Fazil Ahmed, the deputy of Elaziz and the second by Müştak Mayakon, the deputy of Siirt. Ahmed's speech discussed the etymology of the city's name. He argued that its true origin is the Turkish word Elazık, meaning "fertile city", and not the widely held ideas that it was named after the Ottoman Sultan Abdelaziz. After the speech, Atatürk suggested to change the city name to Elazik or simply Elazig".

Pure Turkish words and names

The main sources used to create the list of words are listed below and Figure C.3 displays some example of the sources:

- Besim Atalay, 1935 Türk Büyükleri veya Türk Adları [Turkish Heroes and Turkish Names] . Istanbul: Devlet Basımevi.
- Karauguz, Akin Tahir. 1935. Öz Turk Adları Kilavuzu. Zonguldak: Karaelmas Basımevi.
- Behnan, (Şapolyo) Enver. 1935. Türk Soyadı: 3396 Türk adı [The Turkish Surname: 3396 Turkish Names] Maarif Kitab evi sahibi Tarık. Ankara: Köyhocası Matbaası
- D.K.O. 1935. Öztürkçe Seçme Soy Adları: Karsılıkları ve Manaları. Tefeyyüz Kitaphanesi;
- Orbay, K.Ş. 1935. Öz türkçe Adlar ve Sözcükler: Yeni soy adları [Names and Words in Öztürkçe: The new Surnames]. Istanbul: Hilmi Kitap Evi.
- Vural, M. 1935. Öz türkçe Kadın ve Erkek Adları ve Soy Adları: Öz türkçe Dil Değişimine Armağan [Women's and Men's Proper Names and Surnames in Öztürkçe: A Gift to the Öz türkçe Language Turn] (3. Basılışı) Bursa Bizim Matbaa.
- Osmanlıcadan türkçe ye Cep Kilavuzu [Ottoman-Turkish Pocket Guide] (Istanbul: Devlet Basım Evi, 1935);
- Karauguz, Akın Tahir. 1935. Öz Turk Adları Kilavuzu. Zonguldak: Karaelmas Basımevi ;
- Riza Nour, "Noms propres turcs", Revue de Turcologie 5 (February 1935): 65-72; 65 (circulaire du Ministère de l'Education).
- Öztürkçe seçme soy adları: Müessese Isimleri ve Yeni Adları: Karşılıkları ve manaları, Tefeyyüz Kitaphanesi, 1935

— Ulus Newspapers, available [online](#).⁷

İnönü's Visits

The main source used is the following [website](#), which lists all visits made by İnönü (Yapı Kredi Yayınları, p984), 2016, by Ahmet Demirel, which compiles information from more than a hundred notebook written by İnönü between 1919 and 1973.

People's Houses

I collect and digitize two new sources to obtain the location and year of creation of the People's Houses built between 1932 and 1945 by the single-ruling party, the CHP. The first source is the National Education Statistics for 1944-1945, from the Directorate of Statistics of the Prime Ministry Office.⁸ These records contain information on the name of the city or village where there was a house in 1945, the last year during which houses were built, as well as the number of readers and books. I complement this source with another document from the Prime Ministry Republican Archives in Ankara (BCA) that lists all houses as well as their year of creation. I locate 400 houses. Figure C.5 present the two sources used and Figure C.6 maps the houses as well as the timing of their expansion.

Railroads

I use geo-coded data on train stations collected by [Akgüngör et al. \(2011\)](#) and QGIS software to generate yearly railway shapefiles at a disaggregated level from 1925 to 1949.

Public Primary Schools

I use school administrative censuses, from the Library of the Turkish Statistical Institute (TUIK), in 1925 as well as between 1932 and 1946. The census provide information at the historical district level on the number of schools, teachers, students and graduates. The data is available at the historical district level (approximately 400 districts in 1927, compared to 973 today), that is at a higher level than the rest of my analysis. To obtain harmonized and yearly panel variables, I track all changes in district names as well as their subdivisions, and harmonized the data to the census year 1927, that I then matched to the contemporary district level, my main level of analysis.

Road Network

I digitize historical maps of the road network in 1928, as displayed in Figure C.7, from [Özdemir \(2006\)](#).

7. The list of words starts on [March, the 25th, 1935](#)

8. "Milli Egitim Istatistikleri", Başbakanlık Istatistik Genel Mudurluğu, No. 273, p10-36 (1947).

Turkish Hearth

Data on the Turkish rooms come from a book by Füsün Üstel, *Türk Ocakları 1912-1931* and a book by François Georgeon *Osmanlı-Türk Modernleşmesi 1900-1930*. There is no exact date of creation for all of them. Given this uncertainty, I assume that there were between 50 and 75 Hearths before the visits. This does not affect the results.

Minority Buildings

I use information on the localization of former Armenian and Greek community building (schools or religious building) as of 1912 collected by the [Hrant Dink Foundation's cultural heritage inventory project](#) and available online.

Elites Names and Member of Parliament Biographies

I digitize the biographies of all Turkish deputy members between 1920 and 2010, from the Library of the Grand National Assembly of Turkey (*Türkiye Büyük Millet Meclisi*), in four volumes, are available in pdf format in the following [website](#). The books contain information on the first names and dates of birth of 6,022 deputies, born between 1844 and 1977, as well as the first names of their parents. Figure C.8 provides two examples of biographies used to collect the first names and year of birth of deputies, as well as first names of their parents.

Ancient Trade Roads

I use information on the three ancient trade roads, the Anatolian Silk Road between 1200 and 1400, the Silk Road from the Adriatic between 1200 and 1400 and Ottoman trade roads between 1300 and 1600, made available by the [Old World Trade Routes \(OWTRAD\) Project](#).

Population Data

I also collect additional information on population from the 1914, 1927 and 1935 official population Censuses of the Turkish Republic, from the Turkish Statistical Institute (TUIK). Information is available at the (historical) district level, except for the 1935 census where it is available at the township level (city and villages), for 21,000 towns. For the literacy rate, I use the share of population who is recorded as literate in 1927, available at the district level. I match the historical districts (~ 400) to contemporary districts (~ 973) by tacking subsequent subdivisions and change of names using law decrees and maps used in [Sakallı \(2019\)](#). The total number of births in the birth certificates in 1920 is highly correlated with the 1914 population from the official census, as shown in Figure C.9.

Minority Presence and Villages

To capture historical minority presence at the district level, I use information on the ethnic origin of villages gathered by [Nişanyan \(2010\)](#) and geo-coded by [Sakallı \(2019\)](#). The database provides a list of all localities (towns and villages) whose names were

changed after the creation of the Republic, as well as the linguistic origins of its historic name. I use this information to geo-coded former Armenian, Greek, Arabic and Kurdish villages. I then count the number of villages in each contemporary district, and compute the village density to capture former minority presence.⁹ To run my heterogeneity analysis depending on the presence of minorities locally, I construct an indicator which equals one if the number of Kurdish (or other minorities) villages is above or below the median in a given district.

Occupation during WWI

I use indicators created by [Sakalli \(2019\)](#), that indicates whether an area was occupied or not during the Independence War (1919-1922) by the French, Italian, Greek, British or Russian.

Geographic Covariates

I collect data on geographic covariates : mean annual temperature, precipitation, elevation, growing period length and temperature, and crop suitability indexes for relevant agricultural products in Turkey—wheat, barley, olive, tobacco, potato, cotton. Data are retrieved from the [GAEZ data portal](#) at the grid cell level. I compute the average across cell within contemporary district boundary using QGIS and rescale the resulting average by dividing by 1000.

9. Reassuringly, the number of minority villages is highly correlated with the historical population figures from the 1914 national censuses. I prefer to use the villages as they are easier to aggregate at contemporary district level than the census data.

Figure C.1 – Additional sources: Online websites on various visits

- Afyonkarahisar
- Amasya
- Antalya
- Balıkesir (Ministry of Culture and Tourism)
- Balıkesir
- Burdur
- Bursa, Association of Journalists
- Canakkale
- Canakkale and Gelibolu
- Cankiri
- Devrekani
- Kastamonu
- Dörtöl
- Edremit
- Erzincan
- Eskisehir
- Gaziantep
- Gemlik
- Isparta
- Izmit
- Kayseri
- Manisa
- Menemen
- Narli
- Nigde
- Ordu
- Pertek
- Rize
- Samsun
- Sebinkarahisar
- Tarsus
- Trakya
- Usak
- Cinarcik
- Yalova

Figure C.2 – Example of Historical Newspapers with Information on Atatürk’s Visits



(a) Tokat, Cumhuriyet, April 1930

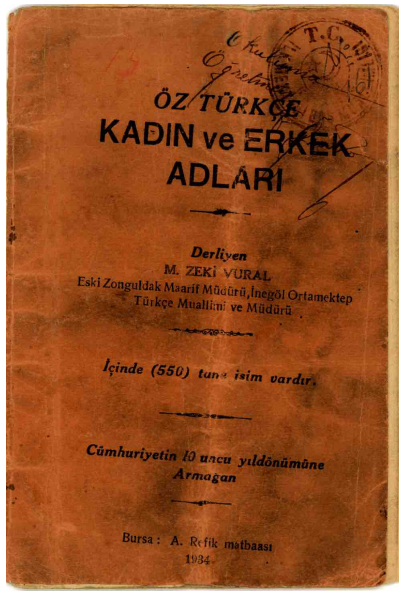
(b) Izmir-Antalya, Milliyet, March 1930

(c) Diyarbakir and Sivas Cumhuriyet, February 1937

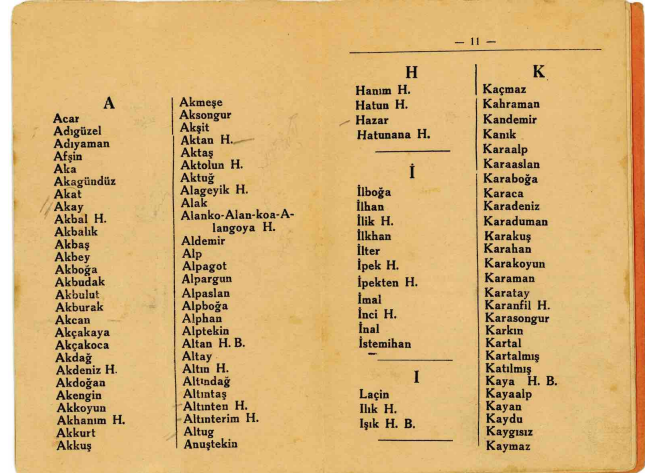
Sources: “A look at History using newspapers”, Istanbul University.

Figure C.3 – Example of sources used to create the list of Pure Turkish words

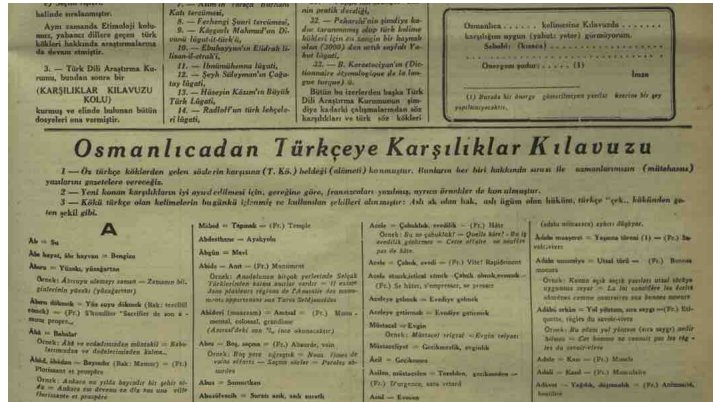
(a) Vural (1935), font page



(b) Vural (1935), p10-11

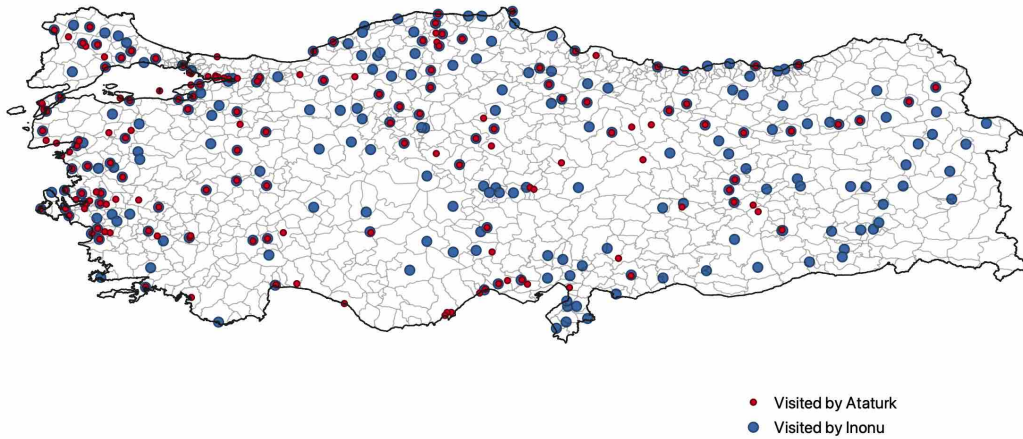


(c) Ulus Newspaper, March 1935



Notes: Examples of a typical book listing the new names and of a newspaper publishing the dictionary from Ottoman to Turkish, used as main sources to create a list of unique words in Pure Turkish to create my outcome variables. *Data Sources: Vural 1935, listed in the Appendix References and Ulus Newspapers, available online.*

Figure C.4 – Map of the Cities Visited by Ataturk and by Inonu



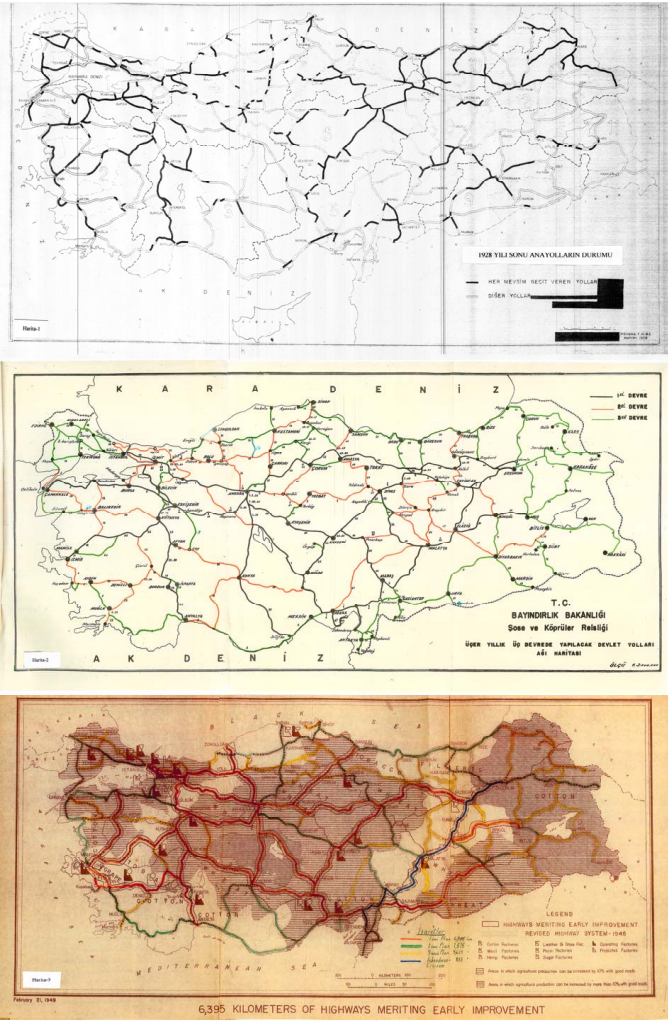
Notes: Author's computations using Kocatürk (2009).

Table A1 – Distribution of the Visits between Atatürk and İnönü

	# of Visits	including	% of Total Visits	including
Ataturk only	49		17%	
Inonu only	140		48%	
Both	105	82 - Ataturk first 15- Inonu first 8- Together	36%	78% - Ataturk first 14% - Inonu first 8% - Together
Total Visits	294		100%	

Notes: Author's computations.

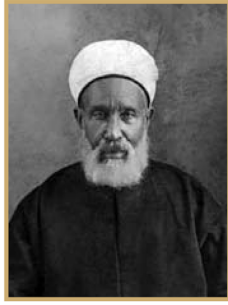
Figure C.7 – Map of the Road Network in 1928



Source: Özdemir (2006)

Figure C.8 – Examples of Biographies of Members of the Grand National Assembly of Turkey

(a)



Abdullah Faik Efendi¹
(Abdullah Çopuroğlu)

Aşağıyirlikler - 1857, İbrahim - Zeynep - Karaisalı Rüştüyesi ve Kayseri Medresesi - Arapça - Müderris - Karaisalı Mahkemesi Üyesi ve Naib Vekili, Karaisalı Müftüsü, Adana Müdâfaa-i Hukuk Cemiyeti Reisi - Osmanlı Medis-i Meb'ûsân I. ve II. Dönem Adana Mebusu – I. Dönem Adana Milletvekili – Evli, 8 Çocuk.
Ölüm Tarihi: 17.07.1939

(b)

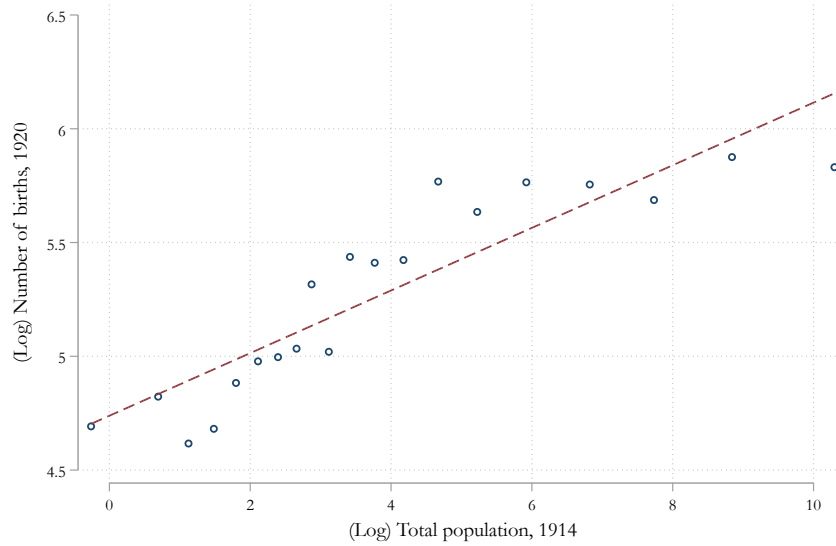


Orhan ERKANLI

Mucur – 1924, Hüseyin Avni – Münevver – Harp Okulu, Harp Akademisi, Türkiye ve Orta Doğu Amme İdaresi Enstitüsü, ABD Zırhlı Birlikler Okulu ve Harp Akademisi – İngilizce – Askerlik, Hariciye – Kurmay Yarbay – Davutpaşa 3. Zırhlı Tugay 3. Tank Tabur Komutanı, Harp Akademisi Öğretmeni, Meksika, Kanada, Yunanistan Büyükelçilikleri Nezdinde Devlet Müşaviri, Hürriyet Gazetesi Genel Yayın Müdürü, Yazar – Millî Birlik Komitesi Üyesi (27.05.1960 – 13.11.1960) ve Genel Sekreteri – Evli, 2 Çocuk.
Ölüm Tarihi: 28.03.1995

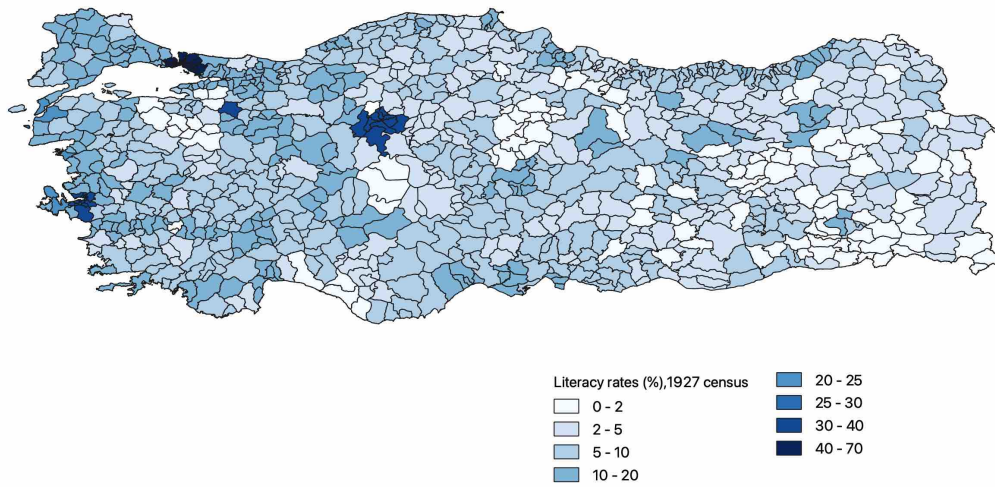
Sources: [TBMM albümü: 1920-2010](#).

Figure C.9 – Correlation between the Number of Birth in 1920 and the total population in 1914



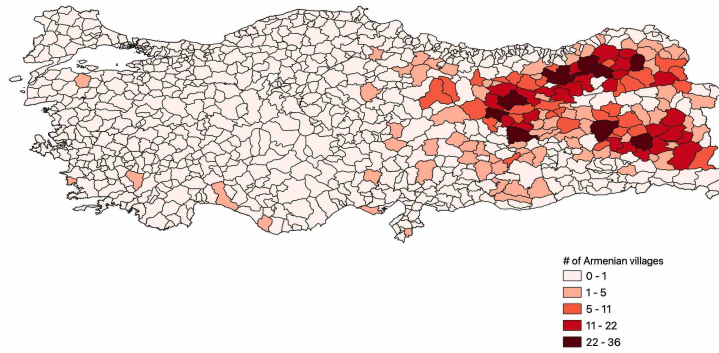
Notes: This figure displays a binscatter between the total number of births in 1920 (using the historical birth certificates database from the Population Office) and the total population in 1914, from the 1914 census, with province fixed effects.

Figure C.10 – Distribution of Literacy Rates in Turkey, 1927 (%)

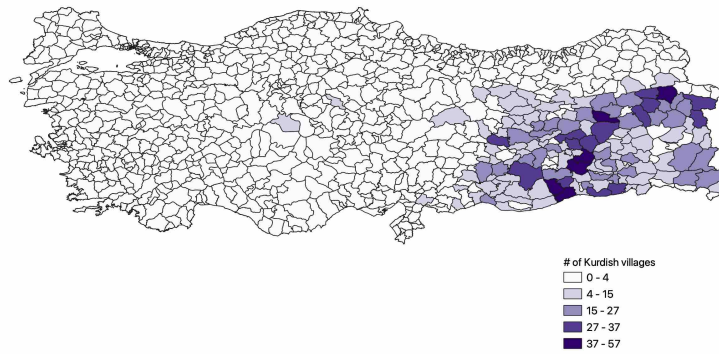


Sources: Turkish National Census, 1927.

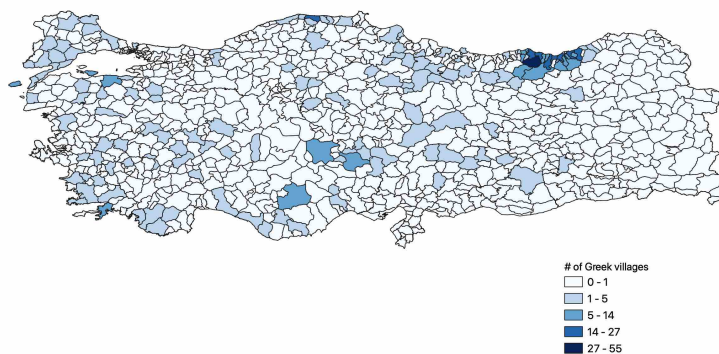
Figure C.11 – Distribution of Armenian, Kurdish and Greek Villages



(a) Armenian Villages



(b) Kurdish Villages



(c) Greek Villages

Note: This figures displays the number of minority villages per contemporary districts, using [Nişanyan \(2010\)](#).

Table A2 – List of visited cities

City name	District	Province	Date of first visit	Total # of Visits	Whether or not it was a target
Ankara	Altindag	Ankara	29Oct1922	10+	1
Eskisehir	Odunpazari	Eskisehir	15Jan1923	10+	1
Arifiye	Arifiye	Sakarya	16Jan1923	1	0
Izmit	Izmit	Kocaeli	16Jan1923	4	1
Bilecik	Bilecik	Bilecik	19Jan1923	2	1
Bursa	Osmangazi	Bursa	20Jan1923	10+	1
Alasehir	Alasehir	Manisa	25Jan1923	1	1
Salihli	Salihli	Manisa	26Jan1923	2	0
Turgutlu	Turgutlu	Manisa	26Jan1923	3	0
Manisa	Sehzadeler	Manisa	26Jan1923	5	0
Menemen	Menemen	Izmir	26Jan1923	5	0
Karsiyaka	Pamukkale	Denizli	27Jan1923	2	0
Karsiyaka	Karsiyaka	Izmir	27Jan1923	2	0
Izmir	Konak	Izmir	27Jan1923	10+	1
Akhisar	Akhisar	Manisa	04Feb1923	3	0
Balikesir	Karesi	Balikesir	06Feb1923	7	1
Balya	Balya	Balikesir	08Feb1923	2	0
Edremit	Edremit	Balikesir	09Feb1923	2	1
Konya	Selcuklu	Konya	14Mar1923	10+	1
Yenice	Tarsus	Mersin	15Mar1923	5	0
Adana	Seyhan	Adana	15Mar1923	7	1
Mersin	Akdeniz	Mersin	17Mar1923	8	1
Tarsus	Tarsus	Mersin	17Mar1923	4	0
Afyonkarahisar	Afyonkarahisar	Afyonkarahisar	23Mar1923	7	1
Kutahya	Kutahya	Kutahya	24Mar1923	2	1
Dumlupinar	Dumlupinar	Kutahya	30Aug1924	1	1
Giresun	Giresun	Giresun	14Sep1924	2	1
Trabzon	Ortahisar	Trabzon	15Sep1924	3	1
Rize	Rize	Rize	17Sep1924	1	1
Ordu	Altinordu	Ordu	19Sep1924	1	1
Samsun	Ilkadim	Samsun	20Sep1924	3	1
Havza	Havza	Samsun	24Sep1924	3	0
Amasya	Amasya	Amasya	24Sep1924	3	0
Turhal	Turhal	Tokat	25Sep1924	3	0
Tokat	Tokat	Tokat	25Sep1924	3	0
Sivas	Sivas	Sivas	27Sep1924	5	1
Zara	Zara	Sivas	28Sep1924	1	0
Hafik	Hafik	Sivas	28Sep1924	1	0
Susehri	Susehri	Sivas	28Sep1924	1	0
Refahiye	Refahiye	Erzincan	29Sep1924	1	0
Erzincan	Erzincan	Erzincan	29Sep1924	1	0
Erzurum	Yakutiye	Erzurum	30Sep1924	1	1
Pasinler	Pasinler	Erzurum	02Oct1924	1	0
Sarikamis	Sarikamis	Kars	04Oct1924	1	0
Kars	Kars	Kars	06Oct1924	1	1
Tercan	Tercan	Erzincan	10Oct1924	1	0
Sebinkarahisar	Sebinkarahisar	Giresun	12Oct1924	1	0
Kayseri	Melikgazi	Kayseri	13Oct1924	4	1
Yozgat	Yozgat	Yozgat	15Oct1924	2	0
Kirsehir	Kirsehir	Kirsehir	17Oct1924	2	1
Dortyol	Dortyol	Hatay	13Jan1925	3	1
Silifke	Silifke	Mersin	27Jan1925	4	0
Tasucu	Silifke	Mersin	28Jan1925	4	1
Tekir	Silifke	Mersin	29Jan1925	2	0
Cankiri	Cankiri	Cankiri	23Aug1925	2	0
Kastamonu	Kastamonu	Kastamonu	23Aug1925	1	1
Seydiler	Seydiler	Kastamonu	25Aug1925	1	0
Kure	Kure	Kastamonu	25Aug1925	2	0
Inebolu	Inebolu	Kastamonu	25Aug1925	1	1
Devrekani	Devrekani	Kastamonu	28Aug1925	1	0
Taskopru	Taskopru	Kastamonu	29Aug1925	1	0
Daday	Daday	Kastamonu	29Aug1925	1	0
Mudanya	Mudanya	Bursa	12Sep1924	10+	1
Istanbul	Fatih	Istanbul	12Sep1924	10+	1
Gemlik	Gemlik	Bursa	04Oct1925	4	0
Soma	Soma	Manisa	10Oct1925	3	0
Kemalpasa	Kemalpasa	Izmir	12Oct1925	2	0
Bornova	Bornova	Izmir	12Oct1925	3	0
Usak	Usak	Usak	16Oct1925	4	0
Bozuyuk	Bozuyuk	Bilecik	19May1926	1	0
Bandirma	Bandirma	Balikesir	13Jun1926	2	1
Urla	Urla	Izmir	30Jun1926	1	0
Cesme	Cesme	Izmir	30Jun1926	1	1
Buyukada	Adalar	Istanbul	14Jul1927	9	0
Tekirdag	Suleymanpasa	Tekirdag	23Aug1928	1	1
Canakkale	Canakkale	Canakkale	01Sep1928	6	1
Eceabat	Eceabat	Canakkale	01Sep1928	1	0
Kucukanafarta	Eceabat	Canakkale	01Sep1928	1	0
Buyukanafarta	Eceabat	Canakkale	01Sep1928	1	0
Conkbayiri	Eceabat	Canakkale	01Sep1928	1	0
Ariburnu	Eceabat	Canakkale	01Sep1928	1	0
Gelibolu	Gelibolu	Canakkale	02Sep1928	1	0
Sinop	Sinop	Sinop	15Sep1928	1	1
Imrali	Karacabey	Bursa	10Aug1929	1	0
Yalova	Yalova	Yalova	19Aug1929	10+	1
Derince	Derince	Kocaeli	15Dec1929	9	0
Isparta	Isparta	Isparta	06Mar1930	1	0
Burdur	Burdur	Burdur	06Mar1930	1	0

Table A2 – List of visited cities

City name	District	Province	Date of first visit	Total # of Visits	Whether or not it was a target
Antalya	Muratpasa	Antalya	06Mar1930	4	1
Serik	Serik	Antalya	09Mar1930	1	0
Gemerek	Gemerek	Sivas	20Nov1930	1	0
Carsamba	Carsamba	Samsun	24Nov1930	1	0
Catalca	Catalca	Istanbul	19Dec1930	1	0
Alpullu	Babaeski	Kirklareli	20Dec1930	1	0
Kirklareli	Kirklareli	Kirklareli	20Dec1930	1	1
Corlu	Corlu	Tekirdag	21Dec1930	3	0
Edirne	Edirne	Edirne	21Dec1930	1	1
Havsa	Havsa	Edirne	25Dec1930	NA	0
Babaeski	Babaeski	Kirklareli	25Dec1930	1	0
Cigli	Cigli	Izmir	27Jan1931	2	0
Selcuk	Selcuk	Izmir	03Feb1931	2	0
Germencik	Germencik	Aydin	03Feb1931	2	0
Aydin	Karacasu	Aydin	05Mar1930	3	1
Nazilli	Nazilli	Aydin	05Mar1930	3	0
Denizli	Merkezefendi	Denizli	05Mar1930	2	1
Malatya	Battalgazi	Malatya	13Feb1931	2	1
Misis	Yuregir	Adana	15Feb1931	2	0
Zonguldak	Zonguldak	Zonguldak	26Aug1931	1	1
Karadenizeregli	Eregli	Zonguldak	26Aug1931	1	0
Cinarcik	Cinarcik	Yalova	22Jul1932	1	0
Sile	Sile	Istanbul	10Oct1932	1	0
Cubuk	Cubuk	Ankara	27Jun1932	2	0
Gaziantep	Sehitkamil	Gaziantep	26Jan1933	1	1
Narli	Pazarcik	Kahramanmaras	27Jan1933	2	0
Fethiye	Fethiye	Mugla	30Jan1933	2	1
Marmaris	Marmaris	Mugla	30Jan1933	2	1
Ahlatlibel	Cankaya	Ankara	05May1933	1	0
Bala	Bala	Ankara	01Feb1934	1	0
Kaman	Kaman	Kirsehir	01Feb1934	1	0
Yerkoy	Yerkoy	Yozgat	20Sep1928	1	1
Sefaati	Sefaati	Yozgat	04Feb1934	1	0
Bogazkoy	Bogazkale	Corum	04Feb1934	1	0
Nigde	Nigde	Nigde	05Feb1934	1	1
Ciftehan	Ulukisla	Nigde	06Feb1934	1	0
Muradiye	Yunusemre	Manisa	09Apr1934	1	0
Foca	Foca	Izmir	09Apr1934	1	0
Gaziemir	Gaziemir	Izmir	10Apr1934	2	0
Kusadasi	Kusadasi	Aydin	10Apr1934	1	0
Seferihisar	Seferihisar	Izmir	11Apr1934	1	0
Bergama	Bergama	Izmir	13Apr1934	1	0
Dikili	Dikili	Izmir	13Apr1934	1	0
Ayvalik	Ayvalik	Balikesir	13Apr1934	1	0
Kucukkuyu	Ayvacic	Canakkale	14Apr1934	1	0
Ezine	Ezine	Canakkale	14Apr1934	1	0
Kizilcahamam	Kizilcahamam	Ankara	16Jul1934	1	0
Gerede	Gerede	Bolu	17Jul1934	1	0
Bolu	Bolu	Bolu	17Jul1934	1	0
Adapazari	Adapazari	Sakarya	02May1931	2	1
Alanya	Alanya	Antalya	18Feb1935	1	0
Muratli	Muratli	Tekirdag	03Jun1936	1	1
Cerkezkooy	Cerkezkooy	Tekirdag	16Aug1937	1	0
Luleburgaz	Luleburgaz	Kirklareli	17Aug1937	1	1
Soke	Soke	Aydin	10Oct1937	1	0
Cetinkaya	Kangal	Sivas	13Nov1937	1	0
Diyarbakir	Baglar	Diyarbakir	15Nov1937	1	1
Elazig	Elazig	Elazig	17Nov1937	1	1
Pertek	Pertek	Tunceli	17Nov1937	1	0
Viransehir	Mezitli	Mersin	21May1938	1	0
Erdek	Erdek	Balikesir	24Jun1938	1	1
Hereke	Korfez	Kocaeli	17Jan1923	1	0
Cerkesli	Dilovasi	Kocaeli	19Jan1923	1	0
Tavsancil	Dilovasi	Kocaeli	19Jan1923	1	0
Gebze	Gebze	Kocaeli	19Jan1923	1	0
Burhaniye	Burhaniye	Balikesir	15Apr1934	1	0
Gomec	Gomec	Balikesir	15Apr1934	1	0
Sivrice	Sivrice	Elazig	17Nov1937	1	0
Maden	Maden	Elazig	17Nov1937	1	0
Korfez	Korfez	Kocaeli	17Jan1923	1	0
Kalecik	Kalecik	Ankara	23Aug1925	1	0
Ilgaz	Ilgaz	Cankiri	23Aug1925	1	0
Ecevit	Kure	Kastamonu	25Aug1925	1	0
Gol	Kastamonu	Kastamonu	30Aug1925	1	0
Kiyik	Kastamonu	Kastamonu	30Aug1925	1	0
Kizilcullu	Buca	Izmir	05Feb1931	1	0
Egirdir	Egirdir	Isparta	06Mar1930	1	0
Talas	Talas	Kayseri	04Feb1934	1	0
Ayvacic	Ayvacic	Canakkale	14Apr1934	1	0
Duzce	Duzce	Duzce	18Jul1934	1	0
Ergani	Ergani	Diyarbakir	15Nov1937	1	0

Note: This table lists all visited cities (and contemporary district and province in which they are), as well as the date of the first visit, the total number of times the city was visited and an indicator of whether it was a targeted city.

Table A3 – Detailed Information on the itineraries with stops along the road

Start Point	Ending Point	Transportation Mode	Departure Date	Arrival Date	# stops	Names of the stops					
Eskisehir	Izmit	Railway	16Jan1923	16Jan1923	6	Arifiye	Hereke	Korfez	Cerkesli	Tavsancil	Gebze
Alasehir	Izmir	Railway	25Jan1923	27Jan1923	5	Salihli	Turgutlu	Manisa	Menemen	Karsiyaka	
Izmir	Balikesir	Railway	04Feb1923	06Feb1923	1	Akhisar					
Balikesir	Izmir	Railway	08Feb1923	10Feb1923	2	Balya	Edremit				
Izmir	Ankara	Railway	18Feb1923	20Feb1923	1	Eskisehir					
Ankara	Adana	Railway	13Mar1923	15Mar1923	2	Konya	Yenice				
Adana	Mersin	Railway	17Mar1923	17Mar1923	1	Tarsus					
Mersin	Ankara	Railway	17Mar1923	25Mar1923	3	Konya	Afyon	Kutahya			
Ankara	Izmir	Railway	31Dec1923	02Jan1924	1	Menemen					
Bursa	Trabzon	Boat	12Sep1924	15Sep1924	1	Mudanya					
Samsun	Sivas	Road	24Sep1924	27Sep1924	3	Havza	Amasya	Tokat			
Sivas	Erzurum	Road	27Sep1924	30Sep1924	4	Zara	Hafik	Susehiri	Erzincan		
Erzurum	Kars	Road	30Sep1924	10Oct1924	1	Sarikamis					
Kars	Sivas	Road	10Oct1924	12Oct1924	3	Tercan	Erzincan	Sebinkarahisar			
Sivas	Ankara	Road	12Oct1924	18Oct1924	3	Kayseri	Yozgat	Kirsehir			
Konya	Adana	Railway	13Jan1925	13Jan1925	1	Dortyol					
Adana	Mersin	Railway	20Jan1925	20Jan1925	4	Yenice	Tarsus	Silifke	Tasucu		
Ankara	Kastamonu	Road	23Aug1925	23Aug1925	1	Cankiri					
Kastamonu	Inebolu	Railway	23Aug1925	25Aug1925	2	Seydiler	Kure				
Inebolu	Kastamonu	Road	25Aug1925	28Aug1925	3	Devrekani	Daday	Taskopru			
Kastamonu	Ankara	Road	29Aug1925	01Sep1925	1	Cankiri					
Ankara	Bursa	Railway	20Sep1925	22Sep1925	2	Izmit	Mudanya	Mudanya	Gemlik		
Balikesir	Izmir	Railway	08Oct1925	11Oct1925	5	Soma	Akhisar	Manisa	Kemalpaşa	Bornova	
Izmir	Konya	Railway	13Oct1925	17Oct1925	1	Usak					
Konya	Ankara	Railway	21Oct1925	22Oct1925	1	Afyon					
Ankara	Mersin	Railway	07May1926	10May1926	4	Konya	Tarsus	Silifke	Tasucu		
Mersin	Adana	Railway	12May1926	16May1926	3	Tekir	Yenice	Dortyol			
Adana	Bursa	Railway	16May1926	20May1926	2	Konya	Bozuyuk				
Bursa	Bandirma	Railway	04Jun1926	13Jun1926	1	Mudanya					
Bandirma	Izmir	Railway	13Jun1926	16Jun1926	4	Balikesir	Manisa	Soma	Menemen		
Izmir	Cesme	Railway	16Jun1926	30Jun1926	1	Urla					
Ankara	Istanbul	Railway	30Jun1927	01Jul1927	1	Izmit					
Istanbul	Bursa	Boat	01Jul1927	15Jul1927	2	Buyukada	Mudanya				
Ankara	Istanbul	Boat	04Jun1928	05Jun1928	1	Buyukada					
Istanbul	Bursa	Railway	23Aug1928	27Aug1928	1	Mudanya					
Istanbul	Canakkale	Boat	01Sep1928	01Sep1928	3	Eceabat	Ariburnu	Anafartalar			
Canakkale	Istanbul	Boat	01Sep1928	Sep1928	1	Gelibolu					
Samsun	Sivas	Road	18Sep1928	18Sep1928	1	Havza	Tokat	Turhal	Amasya		
Kayseri	Ankara	Railway	20Sep1928	21Sep1928	1	Yerkoy					
Ankara	Istanbul	Railway	05Aug1929	06Aug1929	1	Eskisehir					
Istanbul	Bursa	Railway+Road	Aug1929	21Aug1929	2	Yalova	Gemlik				
Yalova	Ankara	Railway	15Dec1929	16Dec1929	1	Derince					
Izmir	Antalya	Railway	05Mar1930	06Mar1930	4	Nazilli	Isparta	Burdur	Serik		
Kayseri	Sivas	Railway	19Nov1930	20Nov1930	1	Gemerek					
Sivas	Samsun	Railway	21Nov1930	21Nov1930	5	Tokat	Turhal	Havza	Amasya	Carsamba	
Istanbul	Kirklareli	Railway	19Dec1930	20Dec1930	2	Catalca	Alpullu				
Kirklareli	Edirne	Railway	21Dec1930	21Dec1930	1	Corlu					
Edirne	Istanbul	Railway	25Dec1930	25Dec1930	2	Havsa	Babaeski				
Istanbul	Bursa	Boat	03Jan1931	04Jan1931	1	Mudanya					
Bursa	Ankara	Railway	05Jan1931	06Jan1931	2	Derince					
Ankara	Izmir	Railway	25Jan1931	27Jan1931	6	Usak	Turgutlu	Cigli	Kemalpaşa	Armutlu	Karsiyaka
Izmir	Aydin	Railway	03Feb1931	03Feb1931	2	Selcuk	Germencik				
Aydin	Denizli	Railway	03Feb1931	04Feb1931	1	Nazilli					
Balikesir	Izmir	Railway	08Feb1931	08Feb1931	1	Cigli					
Izmir	Antalya	Railway	08Feb1931	10Feb1931	1	Tasucu					
Antalya	Malatya	Road	10Feb1931	13Feb1931	3	Tasucu	Tasucu	Mersin			
Dortyol	Adana	Railway	15Feb1931	16Feb1931	1	Yuregir					
Ankara	Istanbul	Boat	20Jul1931	21Jul1931	1	Eskisehir					
Istanbul	Bursa	Railway	Aug1931	Aug1931	2	Mudanya	Yalova				
Istanbul	Zonguldak	Boat	Aug1931	26Aug1931	1	Karadenizeregli					
Istanbul	Ankara	Road	Jun1932	Jun1932	1	Cubuk					
Ankara	Yalova	Boat	15Jul1932	16Jul1932	1	Derince					
Yalova	Istanbul	Boat	16Jul1932	31Jul1932	2	Cinarcik	Buyukada				

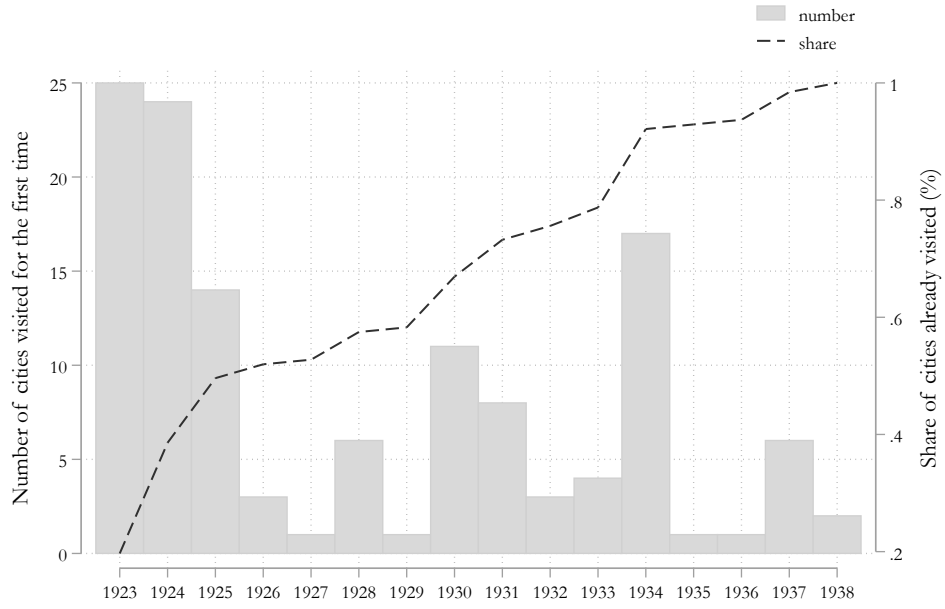
Table A3 – Detailed Information on the itineraries with stops along the road

Start Point	Ending Point	Transportation Mode	Departure Date	Arrival Date	# stops	Names of the stops					
Ankara	Bursa	Railway	16Jan1933	17Jan1933	2	Eskisehir	Derince				
Bursa	Balikesir	Boat + Rail	17Jan1933	21Jan1933	4	Mudanya	Gemlik	Bandirma	Yenikoy		
Kutahya	Adana	Railway	24Jan1933	25Jan1933	3	Afyon	Konya	Yenice			
Gaziantep	Mersin	Railway	27Jan1933	28Jan1933	3	Narli	Adana	Yenice			
Antalya	Izmir	Railway	28Jan1933	31Jan1933	3	Fethiye	Marmaris	Bornova			
Afyonkarahisar	Bursa	Railway	04Feb1933	06Feb1933	1	Bilecik					
Bursa	Istanbul	Railway	06Feb1933	06Feb1933	1	Mudanya					
Ankara	Kirsehir	Railway	01Feb1934	01Feb1934	2	Bala	Kaman				
Kirsehir	Yerkoy	Railway	01Feb1934	02Feb1934	1	Yozgat					
Yerkoy	Kayseri	Railway	04Feb1934	04Feb1934	1	Sefaatti					
Nigde	Konya	Railway	06Feb1934	06Feb1934	1	Ciftehan					
Ankara	Izmir	Railway	07Apr1934	09Apr1934	6	Usak	Salihli	Manisa	Muradiye	Menemen	Foca ;
Izmir	Edremit	Railway	13Apr1934	13Apr1934	5	Bergama	Dikili	Ayvalik	Burhaniye	Gomec	
Edremit	Canakkale	Railway	13Apr1934	15Apr1934	3	Kucukkuyu	Ayvalik	Ezine			
Canakkale	Balikesir	Railway	15Apr1934	15Apr1934	1	Balya					
Balikesir	Ankara	Railway	16Apr1934	17Apr1934	1	Eskisehir					
Istanbul	Yalova	Boat	02May1934	02May1934	1	Bursa					
Yalova	Ankara	Railway	05May1934	06May1934	1	Derince					
Eskisehir	Izmir	Railway	21Jun1934	22Jun1934	5	Afyon	Usak	Turgutlu	Manisa	Gaziemir	
Izmir	Balikesir	Railway	24Jun1934	24Jun1934	3	Akhisar	Soma	Menemen			
Canakkale	Istanbul	Boat	26Jun1934	26Jun1934	1	Buyukada					
Yalova	Ankara	Railway	07Jul1934	08Jul1934	1	Derince					
Ankara	Istanbul	Railway	16Jul1934	19Jul1934	5	Kizilcahamam	Gerede	Bolu	Duzce	Adapazari	
Antalya	Tasucu	Boat	18Feb1935	20Feb1935	1	Alanya					
Tasucu	Mersin	Boat	21Feb1935	21Feb1935	1	Silifke					
Marmaris	Istanbul	Boat	24Feb1935	25Feb1935	1	Canakkale					
Istanbul	Ankara	Railway	03Jun1935	04Jun1935	1	Yalova					
Ankara	Istanbul	Railway	27Jun1935	28Jun1935	2	Bursa					
Istanbul	Muratli	Railway	03Jun1936	03Jun1936	1	Corlu					
Istanbul	Ankara	Railway	08Jun1936	09Jun1936	1	Eskisehir					
Istanbul	Ankara	Railway	28Jul1936	29Jul1936	1	Buyukada					
Konya	Ankara	Railway	08Jan1937	08Jan1937	1	Eskisehir					
Istanbul	Trabzon	Boat	08Jun1937	10Jun1937	1	Yalova					
Trabzon	Istanbul	Boat	12Jun1937	13Jun1937	2	Yalova					
Istanbul	Luleburgaz	Railway	16Aug1937	17Aug1937	2	Cerkezkoy	Corlu				
Istanbul	Ankara	Railway	03Oct1937	04Oct1937	1	Derince					
Ankara	Aydin	Railway	08Oct1937	08Oct1937	2	Nazilli	Soke				
Ankara	Diyarbakir	Railway	12Nov1937	15Nov1937	3	Sivas	Cetinkaya	Malatya			
Diyarbakir	Elazig	Railway	16Nov1937	17Nov1937	3	Maden	Sivrice	Pertek			
Elazig	Adana	Railway	18Nov1937	19Nov1937	1	Yuregir					
Mersin	Ankara	Railway	19Nov1937	20Nov1937	2	Konya	Afyon				
Ankara	Yalova	Railway	20Jan1938	21Jan1938	1	Derince					
Istanbul	Ankara	Road	24Feb1938	25Feb1938	1	Cubuk					
Ankara	Mersin	Railway	20May1938	20May1938	1	Mezitli					
Mersin	Adana	Railway	24May1938	24May1938	1	Tarsus					

Note: This table lists all itineraries where stops were made on the road. The start point and ending points denote the targeted "termini" cities.

C.2 Additional Figures

Figure C.12 – Number and total share of cities visited for the first time, by year



Notes: Author's computations using [Kocatürk \(1988\)](#).

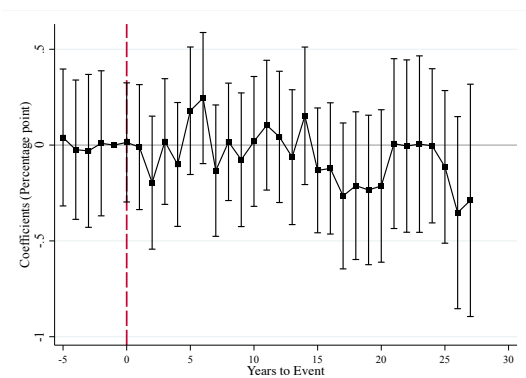
Figure C.13 – Picture of a celebration of the new language at Denizli’s People’s House



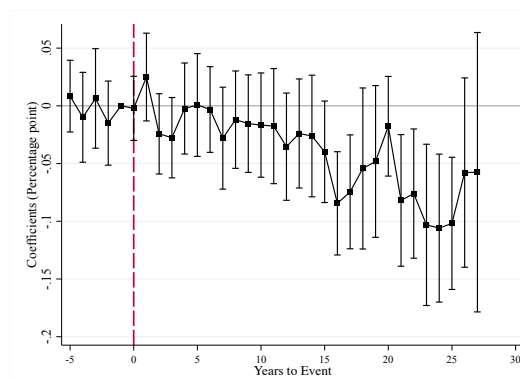
Notes: This picture was taken in front of Denizli’s People’s House on the 26th of September, 1934, the day of the national celebration of the new language (*Dil Bayramı*). On the picture (a), we can see the slogans written on the front of the House “Yurddaş yaban dile yer verme” (“Citizen, do not leave any room to foreign languages”) on the left and “Dilini seven yabancılarla kul olmaz” (“He who loves his language cannot be a slave to foreigners”). Author’s translation to English. Source: (Szurek, 2013, p. 507-510).

Figure C.14 – Effect of the Visits on Other Types of Names

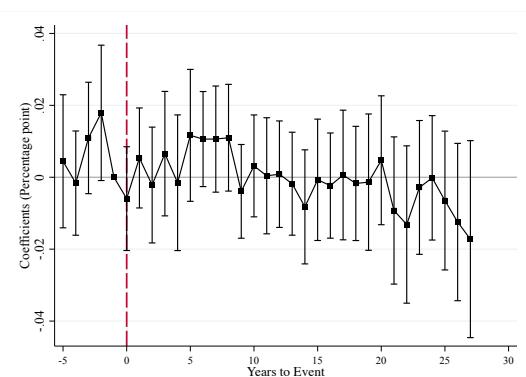
(a) Religious names



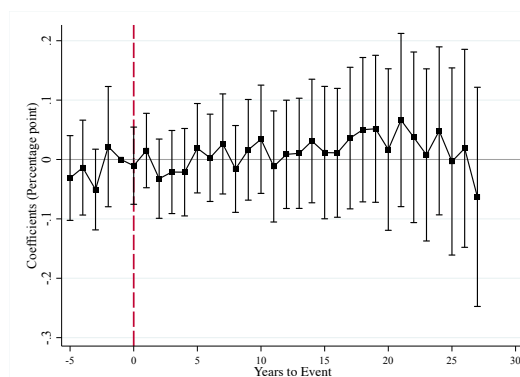
(b) Muhammed



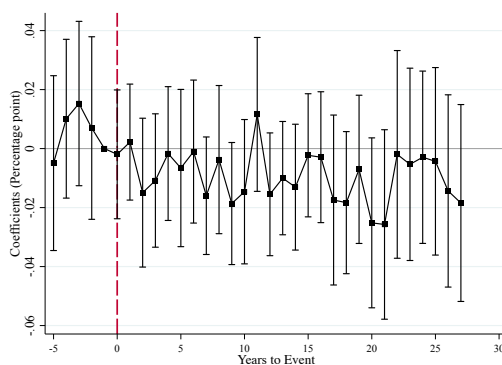
(c) Jewish Names



(d) Kurdish Names

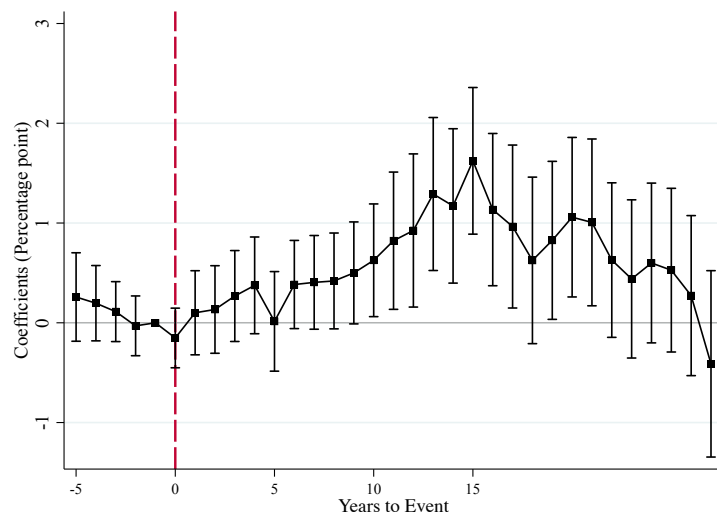


(e) Armenian Names



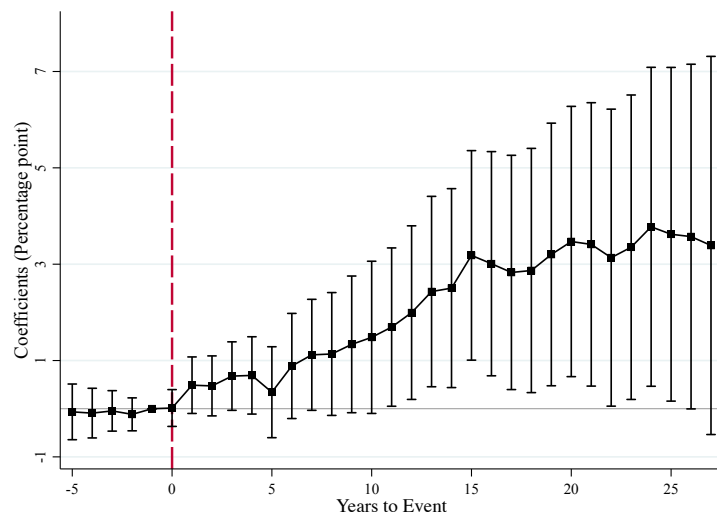
Note: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2 run the full sample. The dependent variables are the share of newborns with religious names (a), named Muhammed (b), with Jewish (c), Kurdish (d) or Armenian (e) names, by district and year. The event is defined as the first time a district is visited by Atatürk. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

Figure C.15 – Robustness check: Event-study results on Pure Turkish names, after re-weighting following Hainmueller (2012)



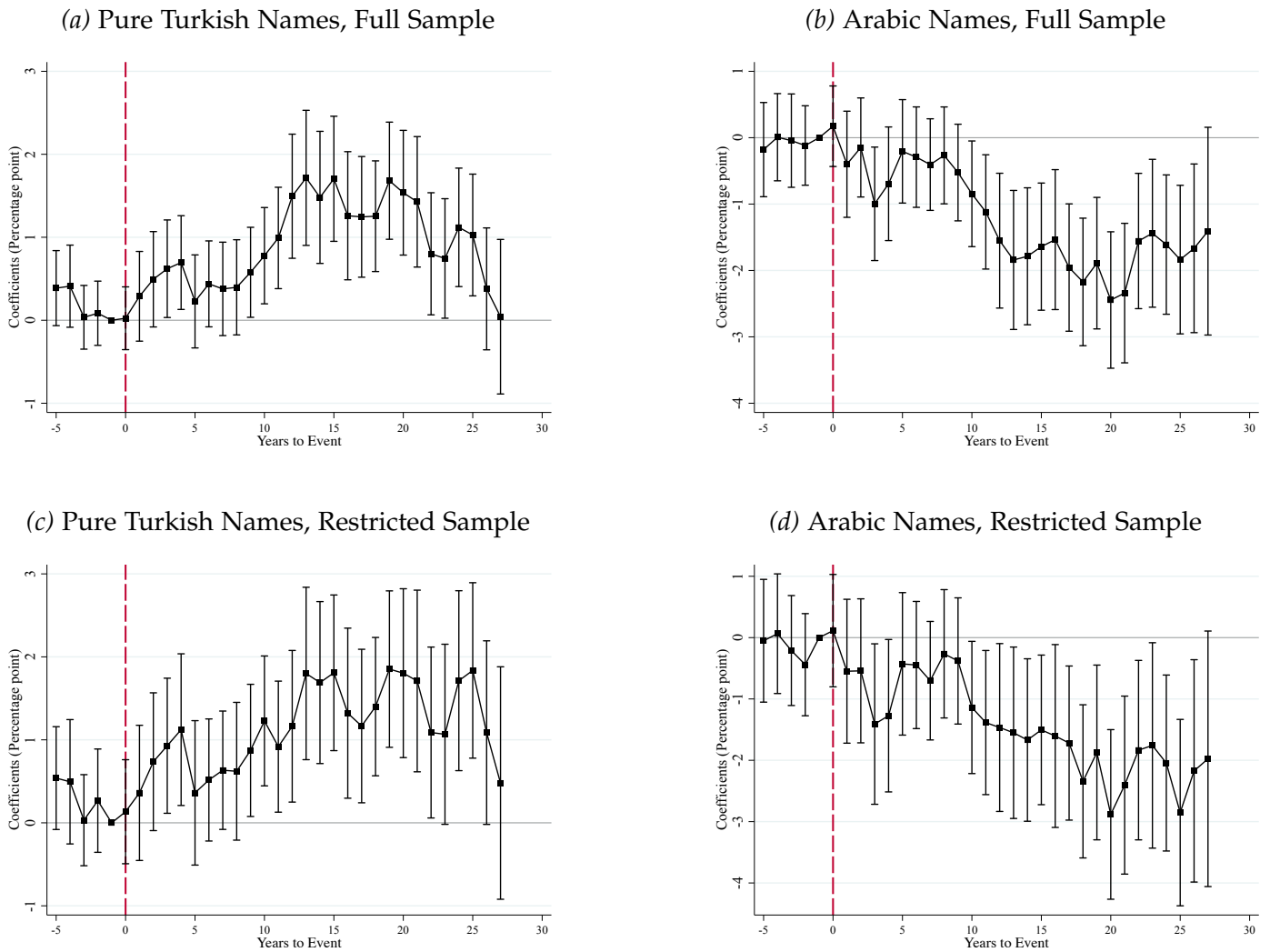
Notes: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2 run the full sample, after re-weighting the observations using entropy balancing following Hainmueller (2012). The dependent variable is the share of newborns with Pure Turkish names in a given district and year. The event is defined as the first time a district is visited by Atatürk. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

Figure C.16 – Robustness Check: Effects on Pure Turkish Names in Visited Districts Only



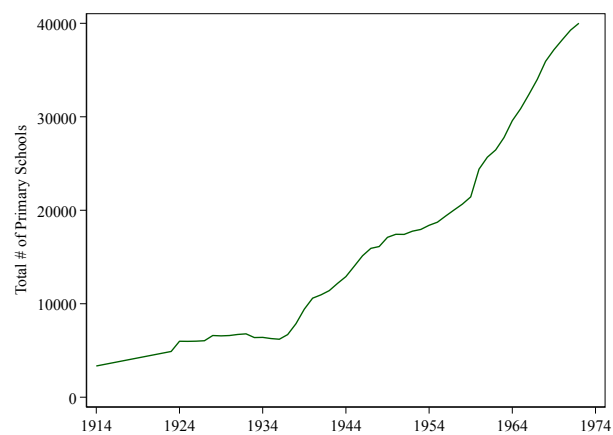
Notes: This figure plots the estimated β_j coefficients from a regression of the form given in equation 3.2 run a sample including only visited districts. The dependent variable is the share of newborns with Pure Turkish names in a given district and year. The event is defined as the first time a district is visited by Atatürk. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

Figure C.17 – Impact of Kemal’s Visit on ‘Pure Turkish’, focusing on children with father born in the same district



Notes: These figures plot the estimated β_j coefficients from a regression of the form given in equation 3.2 run both the full and restricted samples. The dependent variable is the share of newborns with Pure Turkish names in a given district and year, having a father born in the same district. The event is defined as the first time a district is visited by Atatürk. The coefficient of the year prior to the first visit is normalized to zero. The vertical lines reflect the 95% confidence intervals.

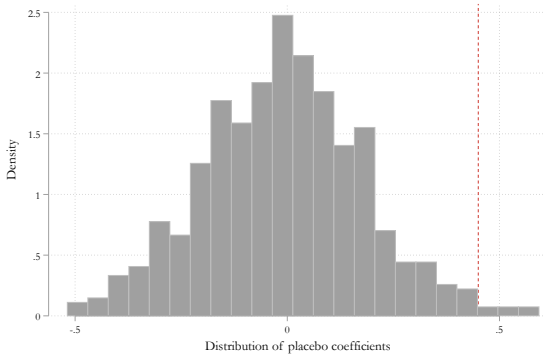
Figure C.18 – Evolution of the number of Primary School in Turkey, 1914-1972



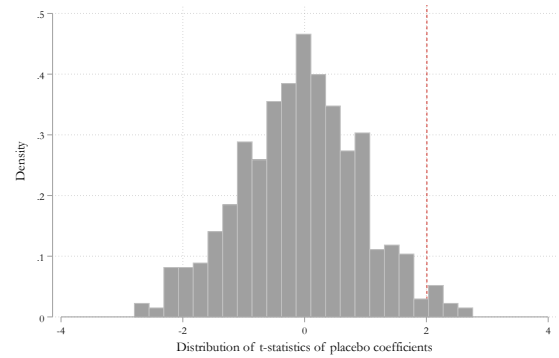
Notes: This figure presents the evolution of the total number of primary schools in Turkey, between 1914 and 1972. Sources: National Education Statistics, from the Directorate of Statistics of the Prime Ministry Office (Maarif ve Milli Egitim Istatistikleri).

Figure C.19 – Placebo Tests

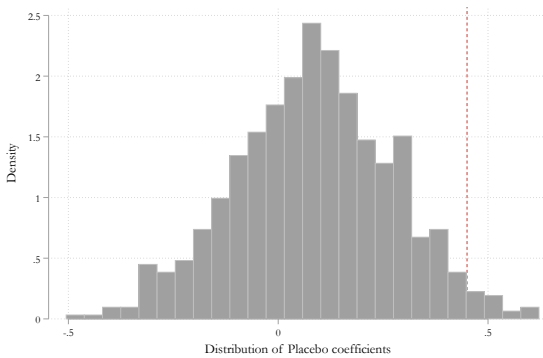
(a) Test 1: Distribution of placebo coefficients



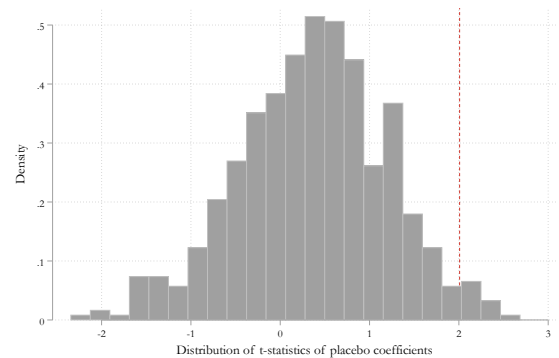
(b) Test 1: Distribution of corresponding t-stats



(c) Test 2: Distribution of placebo coefficients

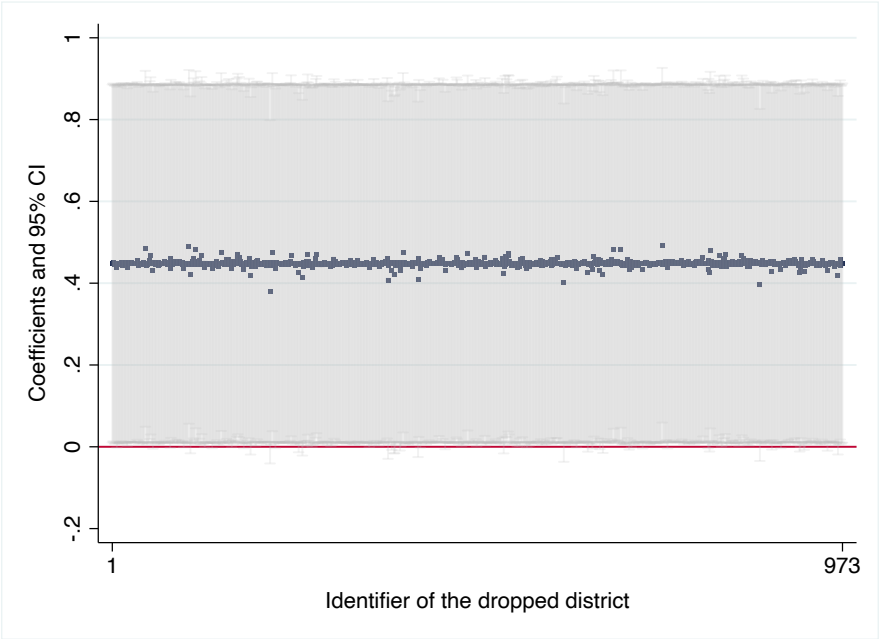


(d) Test 2: Distribution of corresponding t-stats



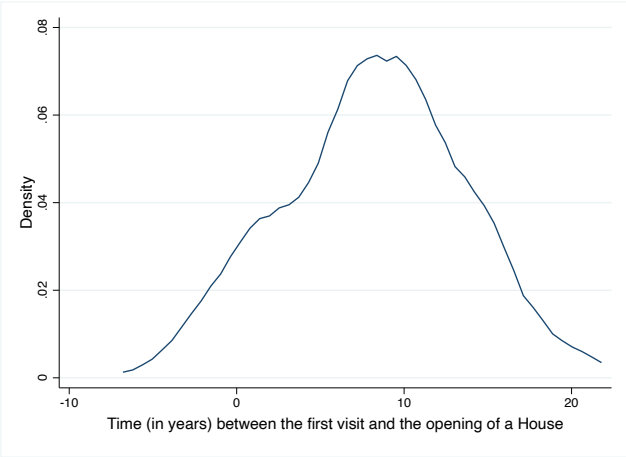
Notes: These figures compare results from the main difference-in-differences specification in equation 3.1, estimated on the full sample to results from two placebo treatments. The first one randomly draws districts and years of treatment, 500 times. Figures (a) and (b) display the coefficients and their t-statistics. The second one randomly assign treatment to non-visited districts within the same province and year when other districts experienced a visit. Figures (c) and (d) displays the corresponding distribution of coefficient and t-statistics. The thick vertical lines indicate the result for the real event for the specification similar to the one presented in Column 1 of Table 3.3. Both the coefficient and its t-statistics from the estimation of the effect of the true event are outside of the corresponding distributions for the placebo events.

Figure C.20 – Sensitivity Analysis: coefficient from the main specification, after dropping one district at a time from the sample



Note: This figure displays results from the main difference-in-differences specification in equation 3.1, estimated on the full sample, removing one district at a time. Each dot plots the corresponding coefficients. The vertical lines reflects the 95% confidence intervals. The estimated coefficients is quite stable and results are not driven by one specific district.

Figure C.21 – Kernel density of the time between the visit and the opening of a People’s House



Notes: I plot the kernel density of the time between the visit and the opening of a People’s House. The sample includes visited districts only. On average, eight years pass between a visit and the opening of a club.

C.3 *Additional Tables*

Table A4 – Summary Statistics of the main database (district level)

	Full sample				Restricted sample			
	Type of locality			Difference (2) - (3)	Type of locality			Difference (6) - (7)
	All (1) Mean (S.D.)	Visited (2) Mean (S.D.)	Non Visited (3) Mean (S.D.)		All (5) Mean (S.D.)	Visited (6) Mean (S.D.)	Non Visited (7) Mean (S.D.)	
				Est. (S.E.)				Est. (S.E.)
Demography and Development								
(Log) Total Nb births, 1920	5.291 (0.839)	5.946 (0.763)	5.169 (0.795)	0.778*** (0.070)	5.631 (0.669)	5.685 (0.642)	5.579 (0.694)	0.107 (0.100)
Has a province center	0.059 (0.235)	0.307 (0.463)	0.012 (0.110)	0.295*** (0.018)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Has a district center	0.412 (0.501)	0.784 (0.428)	0.342 (0.482)	0.442*** (0.042)	0.713 (0.478)	0.667 (0.474)	0.758 (0.479)	-0.092 (0.071)
(Log) Dist. to Railway, 1919	3.897 (1.445)	3.397 (1.706)	3.991 (1.372)	-0.594*** (0.126)	3.316 (1.712)	3.366 (1.715)	3.269 (1.716)	0.098 (0.257)
(Log) Dist. to Road, 1928	2.072 (1.203)	1.539 (1.171)	2.172 (1.183)	-0.633*** (0.104)	1.695 (1.138)	1.604 (1.170)	1.781 (1.106)	-0.178 (0.171)
(Log) Dist. to Ancient Trade Roads	2.362 (1.293)	1.814 (1.351)	2.464 (1.256)	-0.651*** (0.112)	1.992 (1.309)	1.910 (1.268)	2.070 (1.349)	-0.160 (0.196)
City density	0.001 (0.005)	0.003 (0.011)	0.001 (0.002)	0.003*** (0.000)	0.003 (0.004)	0.003 (0.003)	0.003 (0.005)	-0.000 (0.001)
Literacy rate (6+) in 1927 (%)	9.320 (9.502)	12.112 (8.233)	8.790 (9.637)	3.322*** (0.831)	11.214 (10.108)	11.210 (8.190)	11.216 (10.981)	-0.006 (1.298)
Culture and Politics								
(Log) Dist. to Ott. nationalistic clubs	3.859 (0.890)	3.326 (1.070)	3.959 (0.815)	-0.633*** (0.076)	3.576 (0.937)	3.437 (0.906)	3.708 (0.952)	-0.271 (0.139)
Density of Kurdish villages	0.003 (0.006)	0.001 (0.003)	0.003 (0.007)	-0.002** (0.001)	0.001 (0.003)	0.001 (0.003)	0.001 (0.003)	0.000 (0.000)
Density of Arm. villages	0.001 (0.004)	0.001 (0.003)	0.001 (0.004)	-0.001 (0.000)	0.001 (0.002)	0.001 (0.003)	0.000 (0.001)	0.001 (0.000)
Density of Arab villages	0.000 (0.001)	0.000 (0.000)	0.000 (0.001)	-0.000 (0.000)	0.000 (0.001)	0.000 (0.000)	0.000 (0.001)	-0.000 (0.000)
Density of Greek villages	0.003 (0.012)	0.003 (0.011)	0.003 (0.013)	-0.000 (0.001)	0.001 (0.006)	0.001 (0.001)	0.002 (0.008)	-0.001 (0.001)
(Log) Dist. to Istanbul	6.037 (0.969)	5.787 (0.860)	6.083 (0.982)	-0.297*** (0.085)	5.756 (0.828)	5.775 (0.774)	5.738 (0.881)	0.037 (0.125)
(Log) Dist. to Ankara	5.884 (0.645)	5.824 (0.648)	5.896 (0.644)	-0.071 (0.057)	5.785 (0.551)	5.859 (0.587)	5.714 (0.507)	0.145 (0.082)
(Log) Dist. to Border	5.242 (0.953)	5.409 (0.711)	5.210 (0.989)	0.199* (0.084)	5.454 (0.636)	5.438 (0.610)	5.471 (0.663)	-0.033 (0.096)
Occupied after WWI	0.361 (0.481)	0.542 (0.500)	0.327 (0.469)	0.215*** (0.042)	0.511 (0.501)	0.552 (0.500)	0.473 (0.502)	0.079 (0.075)
Density of minority schools	0.015 (0.159)	0.043 (0.292)	0.009 (0.119)	0.034* (0.014)	0.004 (0.008)	0.005 (0.009)	0.003 (0.006)	0.003* (0.001)
Density of minority religious build.	0.047 (0.484)	0.133 (0.853)	0.031 (0.375)	0.102* (0.042)	0.014 (0.028)	0.018 (0.033)	0.010 (0.021)	0.008 (0.004)
Geography								
Coastal	0.079 (0.270)	0.255 (0.437)	0.046 (0.210)	0.209*** (0.023)	0.081 (0.273)	0.174 (0.381)	0.033 (0.180)	0.141*** (0.034)
Ann. precipitation (cm)	0.636 (0.201)	0.645 (0.155)	0.635 (0.209)	0.011 (0.018)	0.603 (0.132)	0.627 (0.121)	0.591 (0.136)	0.036* (0.017)
Mean ann. temperature	11.727 (3.269)	12.600 (3.151)	11.564 (3.267)	1.037*** (0.286)	12.051 (3.029)	12.320 (3.273)	11.913 (2.896)	0.407 (0.388)
Elevation (km)	0.968 (0.547)	0.733 (0.535)	1.012 (0.538)	-0.279*** (0.047)	0.846 (0.519)	0.766 (0.554)	0.887 (0.497)	-0.121 (0.066)
Suitability Index for Cotton	0.046 (0.057)	0.061 (0.059)	0.043 (0.057)	0.018*** (0.005)	0.050 (0.058)	0.060 (0.058)	0.046 (0.057)	0.014 (0.007)
Suitability Index for Olive	0.272 (0.288)	0.403 (0.319)	0.248 (0.275)	0.155*** (0.025)	0.342 (0.312)	0.391 (0.336)	0.318 (0.297)	0.073 (0.040)
Suitability Index for Oat	1.262 (0.156)	1.302 (0.113)	1.254 (0.162)	0.048*** (0.014)	1.282 (0.115)	1.294 (0.108)	1.275 (0.118)	0.019 (0.015)
Suitability Index for Wheat	2.274 (0.329)	2.279 (0.296)	2.273 (0.334)	0.007 (0.029)	2.274 (0.261)	2.257 (0.298)	2.283 (0.240)	-0.026 (0.033)
Suitability Index for Barley	2.279 (0.340)	2.285 (0.302)	2.278 (0.347)	0.007 (0.030)	2.290 (0.263)	2.273 (0.298)	2.298 (0.243)	-0.025 (0.034)
Suitability Index for Tobacco	0.354 (0.173)	0.384 (0.154)	0.348 (0.176)	0.035* (0.015)	0.380 (0.146)	0.352 (0.151)	0.395 (0.141)	-0.043* (0.019)
Suitability Index for Potato	21.133 (8.033)	22.707 (6.854)	20.839 (8.205)	1.868** (0.705)	23.322 (6.989)	22.612 (6.627)	23.685 (7.158)	-1.073 (0.895)
Observations	972	153	819	972	272	92	180	272

Note: This table presents summary statistics for districts that were visited by Atatürk (treatment) and for districts that were not (control), in the full sample (Columns 1 to 4) and in the restricted sample, which excludes departures and final destinations (Columns 5 to 8). Columns 1-3 and 5-7 report means and standard deviations in parentheses. Column 4 and 8 reports differences of group means between Columns 2 and 3 and Columns 6 and 7 respectively, with standard errors in parentheses. The unit of observation is the 2018 Turkish district (973).

Table A5 – Effect of Atatürk’s visits on first names—Piecewise linear regressions

	(1)	(2)	(3)	(4)
	DEP. VAR: SHARE OF NEWBORNS WITH A			
	Pure Turkish Name	Arabic Name	Religious Name	Muhammed
Panel A: Full Sample				
Visited × Post × Between 1-5 years after a visit	0.272 (0.236)	-0.584* (0.330)	-0.085 (0.067)	-0.011 (0.011)
Visited × Post × Between 5-10 years after a visit	0.378* (0.207)	-0.429 (0.300)	-0.022 (0.069)	-0.020 (0.013)
Visited × Post × Between 10-15 years after a visit	0.959*** (0.263)	-1.205*** (0.381)	-0.082 (0.079)	-0.035** (0.017)
Visited × Post × Between 15-20 years after a visit	0.863*** (0.268)	-1.379*** (0.382)	-0.222** (0.096)	-0.059*** (0.019)
Visited × Post × Between 20-28 years after a visit	0.533* (0.291)	-1.080** (0.433)	-0.154 (0.112)	-0.088*** (0.022)
Observations	29760	29760	29760	29760
R-squared	0.750	0.814	0.671	0.444
Mean of outcome at baseline	6.523	71.44	3.386	0.0563
s.d. of outcome	3.432	9.142	2.866	0.269
Panel B: Restricted Sample				
Visited × Post × Between 1-5 years after a visit	0.404 (0.311)	-0.824* (0.426)	-0.042 (0.088)	-0.014 (0.012)
Visited × Post × Between 5-10 years after a visit	0.627** (0.266)	-0.667 (0.406)	-0.021 (0.089)	-0.016 (0.016)
Visited × Post × Between 10-15 years after a visit	1.149*** (0.307)	-1.380*** (0.456)	0.019 (0.099)	-0.023 (0.018)
Visited × Post × Between 15-20 years after a visit	1.190*** (0.338)	-1.758*** (0.499)	-0.226* (0.122)	-0.038 (0.025)
Visited × Post × Between 20-28 years after a visit	0.962** (0.401)	-1.608*** (0.591)	-0.114 (0.144)	-0.062** (0.028)
Observations	8432	8432	8432	8432
R-squared	0.797	0.772	0.482	0.309
Mean of outcome at baseline	6.097	74.38	2.839	0.0383
s.d. of outcome	3.025	6.273	1.953	0.204
Year FE	✓	✓	✓	✓
District FE	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The Table presents the results of piecewise linear regressions, decomposing the effect by time periods. Panel A displays the results using the full sample and Panel B using the restricted sample. The dependent variables are the share of first names by type (Pure Turkish, Arabic, Religious) among newborns. The unit of observation is the district. Baseline controls are described in section 3.4.3. Standard errors in parenthesis, clustered at the district level, the level of the treatment.

Table A6 – Effect of Atatürk’s Visits on Naming Practices, focusing on children with a father born in the same district—Additional effect of a year following a visit

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DEPENDENT VARIABLES: SHARE OF NEWBORNS WITH A:							
	NEW NAMES		TRADITIONAL NAMES		MINORITY NAMES		
	Pure Turkish Name	Arabic Name	Religious Name	Muhammed	Kurdish Name	Armenian Name	Jewish Name
Panel A: Full Sample							
Visited × Years since the Visit	0.041*** (0.015)	-0.071*** (0.021)	-0.012** (0.005)	-0.003*** (0.001)	0.001 (0.003)	-0.001* (0.001)	-0.000 (0.000)
Observations	29760	29760	29760	29760	29760	29760	29760
R-squared	0.698	0.766	0.639	0.445	0.911	0.655	0.804
Panel B: Restricted Sample							
Visited × Years since the Visit	0.054*** (0.019)	-0.083*** (0.027)	-0.007 (0.007)	-0.002** (0.001)	0.001 (0.003)	-0.000 (0.001)	-0.000 (0.000)
Observations	8432	8432	8432	8432	8432	8432	8432
R-squared	0.756	0.719	0.460	0.307	0.804	0.523	0.741
Year FE	✓	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The Table presents the results of the estimation of specification 3.1 but using as treatment variable a variable which equals zero if the district is not visited or not yet visited, and which equals the years since the visits when visited. Panel A displays the results using the full sample and Panel B using the restricted sample. The dependent variables are the share of first names by type (Pure Turkish, Arabic, Religious or minority first names), computed excluding newborns with a father not born in the same district (or for which the place of birth of the father is not available). The unit of observation is the district. Baseline controls are described in section 3.4.3. Standard errors in parenthesis, clustered at the district level, the level of the treatment. Following a visit, in the restricted sample, the average increase in “Pure Turkish” names among newborns with fathers born in the same district equals 0.05 percentage points for any year following a visit.

Table A7 – Balance in district-level characteristics, following Hainmueller (2012)

	Treated group		Control group					
	Mean	Var.	Before weighting			After weighting		
			Mean	Var.	Difference	Mean	Var.	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Demography and Development								
(Log) Total Nb births, 1920	5.95	.58	5.17	.63	1.02	5.93	.61	.02
Has a province center	.31	.21	.01	.01	.64	.3	.21	.02
Has a district center	.78	.18	.34	.23	1.03	.78	.25	.02
(Log) Dist. to Railway, 1919	3.4	2.91	3.99	1.88	-.35	3.41	3.13	-.01
(Log) Dist. to Road 1928	1.54	1.37	2.17	1.4	-.54	1.55	1.96	-.01
(Log) Dist. to Ancient Trade Roads	1.81	1.82	2.46	1.58	-.48	1.83	1.63	-.01
City density	0	0	0	0	.25	0	0	0
Culture and Politics								
(Log) Dist. to Ottoman nationalistic clubs	3.33	1.14	3.96	.67	-.59	3.34	1.59	-.01
Density of Kurdish villages	0	0	0	0	-.61	0	0	-.02
Density of Arm. villages	0	0	0	0	-.23	0	0	0
Density of Greek villages	0	0	0	0	0	0	0	0
(Log) Dist. to Istanbul	5.79	.74	6.08	.96	-.35	5.79	1.22	0
(Log) Dist. to Ankara	5.82	.42	5.9	.41	-.11	5.83	.45	-.01
(Log) Dist. to Border	5.41	.51	5.21	.98	.28	5.4	.68	.01
Occupied after WWI	.54	.25	.33	.22	.43	.54	.25	.01
Density of minority schools	.04	.09	.01	.01	.12	.04	.1	0
Density of religious minority buildings	.13	.73	.03	.14	.12	.13	.97	0
Geography								
Coastal	.25	.19	.05	.04	.48	.25	.19	.01
Ann. precipitation (cm)	645	23	634	43	.07	646	21	-.01
Mean ann. temperature	12	9	11.56	10.67	.33	12.58	8.27	.01
Elevation (km)	732	286	1011	289	-.52	737	229	-.01
Suitability Index for cotton	.06	0	.04	0	.32	.06	0	0
Suitability Index for olive	.4	.1	.25	.08	.49	.4	.08	.01
Suitability Index for oat	1.3	.01	1.25	.03	.42	1.3	.02	0
Suitability Index for wheat	2.28	.09	2.27	.11	.02	2.28	.07	0
Suitability Index for barley	2.29	.09	2.28	.12	.02	2.29	.09	0
Suitability Index for tobacco	.38	.02	.35	.03	.23	.38	.02	0
Suitability Index for potato	22.71	46.97	20.84	67.33	.27	22.63	81.25	.01

Notes: This table presents the mean and variance of historical and geographic covariates in visited districts (Columns 1 and 2) and non-visited districts (Columns 3 and 4), and the differences of the groups means between Columns 1 and 3, before re-weighting. Columns 6-7 presents the mean and variance in the control group, after re-weighting with the formula of [Hainmueller \(2012\)](#). The unit of observation is the 2018 Turkish district (n=973).

Table A8 – Effect of Atatürk’s Visits - Heterogeneity analyses depending on minority presence

<i>Dep. Var: Share of newborns with a</i>	(1)	(2)	(3)	(4)	(5)
		"PURE TURKISH" NAME			
Panel A: Full Sample					
Visited × Post	0.44** (0.20)	0.44** (0.22)	0.44** (0.22)	0.44** (0.22)	0.44** (0.22)
Visited × Post × Number of Arm. villages		-0.00 (0.03)			
Visited × Post × Number of Greek villages			-0.00 (0.04)		
Visited × Post × Number of Kurdish villages				-0.00 (0.03)	
Visited × Post × Number of Minority villages					-0.00 (0.02)
Observations	29760	29760	29760	29760	29760
R-squared	0.750	0.750	0.750	0.750	0.750
Panel B: Restricted Sample					
Visited × Post	0.56** (0.26)	0.57** (0.28)	0.67** (0.33)	0.64** (0.28)	0.64** (0.28)
Visited × Post × Number of Arm. villages		-0.02 (0.03)			
Visited × Post × Number of Greek villages			-0.16 (0.19)		
Visited × Post × Number of Kurdish villages				-0.07** (0.03)	
Visited × Post × Number of Minority villages					-0.04* (0.02)
Observations	8432	8432	8432	8432	8432
R-squared	0.796	0.796	0.796	0.796	0.796
Year FE	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This Table presents the results of the estimation of specification 3.1 on the full sample (Panel A) and on the restricted sample (Panel B), adding as heterogeneity variable the total number of minority villages, by type, in a given district. The dependent variables are the share of Pure Turkish name. Baseline controls included. Standard errors in parenthesis, clustered at the district level, the level of the treatment.

Table A9 – Effect of Atatürk’s Visits - Heterogeneity analyses depending on literacy rates and distance to a Turkish Hearth

	(1)	(2)	(3)	(4)	(5)
<i>Dep. Var: Share of newborns with a</i>					
		"PURE TURKISH" NAME			
Panel A: Full Sample					
Visited × Post	0.44** (0.20)	-0.25 (0.32)	-0.43 (0.40)	0.12 (0.24)	0.62** (0.29)
Visited × Post × Literacy rate (6+) in 1927 (%)		0.06** (0.02)			
Visited × Post × Male literacy rate (6+) in 1927 (%)			0.05** (0.02)		
Visited × Post × Female literacy rate (6+) in 1927 (%)				0.07** (0.03)	
Visited × Post × Dist. to Ottoman nationalistic club					-0.00 (0.01)
Observations	29760	29760	29760	29760	29760
R-squared	0.750	0.750	0.750	0.750	0.750
Panel B: Restricted Sample					
Visited × Post	0.56** (0.26)	-0.51 (0.35)	-0.88** (0.40)	0.07 (0.28)	1.15*** (0.35)
Visited × Post × Literacy rate (6+) in 1927 (%)		0.10*** (0.03)			
Visited × Post × Male literacy rate (6+) in 1927 (%)			0.08*** (0.02)		
Visited × Post × Female literacy rate (6+) in 1927 (%)				0.10*** (0.04)	
Visited × Post × Dist. to Ottoman nationalistic club					-0.01*** (0.01)
Observations	8432	8432	8432	8432	8432
R-squared	0.796	0.796	0.796	0.796	0.796
Year FE	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This Table presents the results of the estimation of specification 3.1 on the full sample (Panel A) and on the restricted sample (Panel B), adding as heterogeneity variable the literacy rates, by gender, in a given district and the distance to the closest Turkish Hearth. The dependent variables are the share of Pure Turkish name. Baseline controls included. Standard errors in parenthesis, clustered at the district level, the level of the treatment.

Table A10 – The Visits are among the main predictors of the Opening of a People’s House

<i>Dep. Variables: Years of opening Model</i>	(1) (Log) Dist. House 1932-1938 OLS	(2) Has a House 1932-1938 Probit	(3) (Log) Dist. House 1939-1944 OLS	(4) Has a House 1939-1944 Probit
(Log) Dist. to the nearest Visit	0.10*** (0.01)	-2.27*** (0.40)	-0.01 (0.01)	0.16 (0.18)
(Log) Dist. nearest to the nearest Province center	0.65*** (0.02)	1.88 (1.18)	-0.03* (0.02)	-1.60*** (0.41)
(Log) Dist. nearest to the District Province center	0.40*** (0.01)	-5.69*** (0.46)	0.25*** (0.01)	-2.10*** (0.10)
(Log) Dist. to the road, 1928	0.03*** (0.00)	-0.34*** (0.13)	0.03*** (0.00)	-0.22*** (0.07)
(Log) Dist. to the nearest Ott. nationalistic club	0.14*** (0.01)	-1.07*** (0.39)	0.01 (0.01)	1.27*** (0.21)
(Log) Dist. to Ankara	-0.52*** (0.05)	-0.85 (4.98)	-0.10* (0.05)	-0.14 (1.47)
(Log) Dist. to the nearest rebellion	0.01 (0.01)	-1.01*** (0.35)	-0.06*** (0.01)	-0.69*** (0.21)
(Log) Dist. to the nearest battle	0.10*** (0.01)	-1.01*** (0.32)	-0.04*** (0.01)	-0.82*** (0.19)
Nb of arab villages within 10 km	0.01*** (0.00)	0.81*** (0.28)	0.00 (0.00)	0.10 (0.07)
Nb of arab villages within 20 km	0.00** (0.00)	-0.31** (0.15)	0.01*** (0.00)	0.02 (0.04)
Nb of kurd villages within 20 km	-0.00*** (0.00)	-0.12*** (0.03)	-0.01*** (0.00)	-0.02 (0.02)
(Log) Dist. to the nearest Arab village	-0.09*** (0.03)	5.76*** (1.70)	-0.48*** (0.03)	-1.13 (0.88)
(Log) Dist. to the nearest Arm. village	-0.13*** (0.04)	-7.63*** (2.71)	-0.50*** (0.04)	-3.59* (1.87)
(Log) Dist. to the nearest Greek village	0.02 (0.04)	-2.75 (3.05)	-0.30*** (0.05)	3.35*** (1.04)
Growing Period Length	-0.00** (0.00)	0.02*** (0.00)	-0.00*** (0.00)	0.00* (0.00)
Elevation	0.00*** (0.00)	-0.00* (0.00)	0.00** (0.00)	-0.00*** (0.00)
Suitability index for oat	0.00*** (0.00)	-0.00 (0.01)	-0.00*** (0.00)	0.00 (0.00)
Suitability index for olive	-0.00*** (0.00)	-0.01 (0.01)	0.00*** (0.00)	-0.01*** (0.00)
Constant	1.73*** (0.21)	43.07** (17.45)	6.94*** (0.24)	9.81 (8.31)
Observations	35,703	11,248	35,703	7,105
R-squared	0.859		0.833	
District FE	✓	✓	✓	✓
Controls selected by LASSO	✓	✓	✓	✓

Notes: *** p<0.01, ** p<0.05, * p<0.1. This table shows the effect of the logarithm of the distance to the nearest visited on the distance to the nearest house (Columns 1 and 3) and on the probability on having a house (Columns 2 and 4) for houses opened between 1932-1938 (Columns 1 and 2) and 1939-1944 (Columns 3 and 4). Regression models include district fixed effects and covariates selected by a LASSO procedure. Robust standard errors in parentheses.

Table A11 – Visits are predictors of the Opening of the Houses: Robustness to Spatial Autocorrelation

Dependent Variable: (Log.) Distance to the nearest People’s Houses		
Assumption about variance-covariance matrix:		(Log.) Dist. Nearest visited city
Coefficient		0.08
1	Baseline: OLS Regression with district fixed effects	(0.01)***
	Conley correction for spatial correlation within:	
2	10 km	(0.02)***
3	20 km	(0.03)**
4	50 km	(0.04)*
5	100 km	(0.04)*
Observations		35,703
Controls Selected by lasso		✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The table shows that the results on the prediction of the opening of a People’s House, presented in Table 3.5 are robust to adjusting the standard errors to spatial correlation at 10, 20, 50 and 100km. The coefficient and standard errors of the distance to the nearest visit at baseline are the one presented in Column (2), Table 3.5.

Table A12 – Effect of the visits and of the expansion of railway

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Dep. Var:</i>	SHARE OF NEWBORNS WITH A:											
	Pure Turkish Name			Arabic Name			Kurdish Name			Religious Name		
Has a Railway × Post	0.300* (0.175)	0.278 (0.174)	0.317* (0.190)	-0.395 (0.261)	-0.363 (0.261)	-0.446 (0.291)	-0.006 (0.066)	-0.006 (0.066)	-0.017 (0.077)	-0.060 (0.080)	-0.058 (0.080)	-0.042 (0.090)
Visited × Post		0.346* (0.206)			-0.513* (0.291)			-0.001 (0.042)			-0.039 (0.067)	
Has a Railway × Post × Visited Before			-0.115 (0.433)			0.333 (0.575)			0.076 (0.110)			-0.124 (0.150)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760	29760
Mean of outcome	6.523	6.523	6.523	71.44	71.44	71.44	1.467	1.467	1.467	3.386	3.386	3.386
s.d. of outcome	3.432	3.432	3.432	9.142	9.142	9.142	3.178	3.178	3.178	2.866	2.866	2.866

Notes: *** p<0.01, ** p<0.05, * p<0.1. This table presents the results of the estimation of specification similar to equation 3.1, but where the treatment variable is a dummy variable switching to one the first year a railway line is opened in a given district. The dependent variables are the share of first names by type (Pure Turkish, Arabic, Religious or minority first names). The unit of observation is the district. Baseline controls included. Standard errors in parenthesis, clustered at the district level, the level of the treatment.