IMT Institute for Advanced Studies, Lucca

Lucca, Italy

Essays on Political Economy

PhD Program in Economics, Markets and Institution

XXVI Cycle

By

Alessandro Belmonte

2014

The dissertation of Alessandro Belmonte is approved.

Program Coordinator: Prof. Davide Ticchi,
IMT Institute for advanced studies Lucca

Supervisor: Prof. Andrea Vindigni,
IMT Institute for advanced studies Lucca

The dissertation of Alessandro Belmonte has been reviewed by:

Prof. Francesco Passarelli,
University of Teramo and Bocconi University

Prof. Edoardo Grillo,
Collegio Carlo Alberto

IMT Institute for Advanced Studies, Lucca

2014

Contents

List of Figures

Acknowledgements

Chapter 2 is the main paper included in the thesis. I started to work at it in Summer 2012 and previous preliminary versions have been presented at the 2014 Political Economy Workshop at Catholic University of Milan and at IMT Institute for Advanced Studies Lucca in a talk that I held in June 2014. I am deeply in debt to my Advisor Prof. Andrea Vindigni who introduces me to the beauties of Political Economy. He encourages me a lot and his suggestions contribute to improve definitely the paper. I am especially grateful to Prof. Davide Ticchi for his constant support and useful advices. In the last two years I had the opportunity to talk with many persons about this work. Among others, I would like to thank Prof. Roland Benabou, Prof. Patrick Francois, Prof. Francesco Passarelli, and Prof. Gilles Saint-Paul for extremely helpful comments and suggestions.

Chapter 4 is the reproduction of the article "The Italian primary school-size distribution and the city-size: a complex nexus", *(Nature) Scientific Reports* (2014). The paper is coauthored by Riccardo Di Clemente and Prof. Sergey V. Buldyrev (Yeshiva University, USA). The paper required two years during which I had time to know very well Riccardo whom I would like to especially thank for his contribution, dedication and friendship. I am really grateful to Prof. Sergey V. Buldyrev, who despite the geographical distance among us, believed in the project, supported it and stimulated our research. The paper has been additionally presented in several conferences: *ECCS' 13* in Barcelona, *SigmaPhi2014* International Conference on Statistical Physics in Rhodes, Greece, in June 2014, and *ECCS' 14* in Lucca. The paper has received a great deal of media coverage in September 2014 both in the

Italian scientific community and in the most important Italian national-wide newspapers such as *La Stampa*.

Vita

Jan. 19, 1984 Born, Palermo (Palermo), Italy

Current Appointments

2011 Ph.D. Candidate in Economics, Market and Institution at IMT-Lucca Institute for Advanced Studies XXVI ciclo Supervised by Prof. Andrea Vindigni.

2014 Research fellow (Ricercatore di III Livello) at INVALSI, Istituto Nazionale per la VALutazione del SIstema educativo e di formazione, in econometrics.

Past appointments

2013/14 Research assistant to Professors Roland Bénabou, Davide Ticchi, and Andrea Vindigni within the project 'The Politics of Science, Religion, and Growth'.

2010/11 Collaboration with the Italian Antitrust Authority (Autoritá Garante della Concorrenza e del Mercato, AGCM) . Evaluation of merger cases in pharmaceutics and transport sectors.

2010 Collaboration with the Italian Ministry of Economics and Finance, Ragioneria Generale dello Stato - Servizio Studi Dipartimentale - Office VII. Implementation of a simulation model to forecast education demand and teacher demand by italian provinces

Research visit

2014 Visiting Research Student at the Department of Economics, University of Warwick.

2013 Visiting Research Student at the Paris School of Economics, PSE.

Higher Education

2010 M.Sc. in European Lawyer – Market: Competition and Consumers, Università degli Studi di Roma Tre.
Admitted with tuition fee waiver granted by the Italian Antitrust Authority.

2010 M.Sc. in Economics (with Honors), Università degli Studi di Modena e Reggio Emilia.
Thesis: *"The exploitation of common goods. Analysis of a model of congestion"*.
Supervisor: Prof. Luigi Brighi.

2006 B.Sc. in Economics of Development (with Honors), Università degli Studi di Palermo.

Ph.D. Schools

2011 CIdE XXII Course of Econometrics for PhD students Summer School in Econometrics 2011.
CIdE - Centro Interuniversitario di Econometria (Bertinoro, Italy).

2010 SSE - Scuola Superiore di Economia, Università Cà Foscari di Venezia e Università di Padova (SDEM)
Spring School in Economics: Theoretical and Empirical Challenges in Economics

Honors

2013 *"Talent At Work"* EU Lifelong Learning Programme 5-months grant 2013.

2011 Admission to the Ph.D. program in Economics, Markets and Institution relations at IMT Institute for Advanced Studies Lucca, with 3-years full scholarship, tuition fee waiver and free lodging.

2010 Admission to the M.Sc. in European Lawyer – Market: Competition and Consumers, Università degli Studi di Roma Tre, granted by the Italian Antitrust Authority (AGCM) Scholarship

2010 M.Sc degree with Honors

2007 B.Sc degree with Honors

Publications

1. Belmonte, A., Di Clemente, R. & Buldyrev, S. V. The italian primary school-size distribution and the city-size: a complex nexus *(Nature) Scientific Reports* **4**, 5301 (2014) DOI :10.1038/srep05301
2. Belmonte A., Pennisi A., Education reforms and teachers needs: a long-term territorial analysis *Italian Journal of Regional Science - IJRS* **1**, 12: 87-114 (2013) DOI :10.3280/SCRE2013-001005
3. Belmonte A., Pennisi A., L'evoluzione della domanda scolastica a livello provinciale *Rivista trimestrale di scienze dell'amministrazione scolastica* **3**: 5-13 (2011)

Conference

Poster

1. Belmonte, A., Di Clemente, R., & Buldyrev, S. V. A Complexity Approach To The Italian Primary School-Size Distribution *ECCS' 13 Book of Abstract: European Conference on Complex Systems* (Barcelona, Spain 2013)

Talks

1. Belmonte, A. Sophisticated Electoral Accountability. A Political Psychology Agency Theory *IMT Institute for Advanced Studies Lucca* (Lucca, Italy 2014).
2. Belmonte, A. Sophisticated Electoral Accountability. A Political

Psychology Agency Theory *2014 Political Economy Workshop, Catholic University* (Milan, Italy 2014).

3. Belmonte, A., Di Clemente, R., & Buldyrev, S. V. On the spatial distribution of the Italian primary school-size *SigmaPhi2014 International Conference on Statistical Physics* (Rhodes, Greece 2014).

4. Belmonte, A., Di Clemente, R., & Buldyrev, S. V. On the spatial distribution of the Italian primary school-size *ECCS'* 14 *European Conference on Complex Systems* (Lucca, Italy 2014).

5. Belmonte, A., & Pennisi A., Un modello di simulazione di medio-lungo termine del fabbisogno scolastico. Simulazioni 2010-2027 ed impatti delle riforme in corso *Italian Ministry of Economy and Finance* (Rome, Italy 2010).

Abstract

While a vast literature has been collected pointing out the role of the human capital on economic growth, a few has been said, in Economics, on the mechanism through which education directly impacts on democracy. The present dissertation proposes a novel microfoundation of this nexus based on the paramount role of education as economic growth engine and determinant of political participation. The first two works introduce elements of political psychology in order to shed lights on individual cognitive process that might favor, overall, a culture of democracy. Education is then a cognitive tool that citizens/voters can use to decode the information content of political signals and to keep rulers in charge accountable. We formally show that the entire initial distribution of education matters for a successful democracy and that the median is pivotal in the political process. Motivated by that in the last work we propose a statistical analysis of the distribution of the Italian primary school service. Primary schools tend to distribute in a complex way according to geographical features of the territory, schooling aged population density, and possible interactions between the two. Despite the school system is financed at a State level, we outline the persistence of remarkable differences not directly attributable to historical divergences among different macro-area of the country but rather between montane areas and more dynamic regions deputed to explain economic and political divergences.

Chapter 2

We propose a political agency model where rent-maximizer rulers are constrained by sophisticated principals/producers that use an awareness-management model à la Bénabou and

Tirole. Sophistication is explained by educational attainments and producers are endowed with different levels of education, that increase over time with human capital investments. We allow education to be both the engine of growth and a determinant of political participation; in equilibrium, more educated societies are more able to punish politicians that, in turn, invest more in productive public goods such as infrastructure, roads or legal rules for contracts enforcement. We prove the existence of multiple steady states featuring, respectively, a sophisticated society with congruent politicians in office, and a naive society ruled by dissonant politicians. Finally, we address inequality concerns and show how, for intermediate values, inequality opposingly hits citizens and ruler and only the latter is found to better off; conversely, citizens are averse to inequality, contributing to explain, via sophisticated accountability, why most people dislike living in a society which is too unequal.

Chapter 3

The paper originally attempts to explain the rise of the new wave of populism in Europe and the persistence of the Latin American populism. Such phenomena rose an unresolved political puzzle according to which populist politicians has been widely supported by the electorate while ultimately hurt the economic interests of the majorities. We address this puzzle by looking at the electorate side and, specifically, at individual citizens that are endowed with different level of political sophistication. According to the Political Psychology literature, we approximate political sophistication in terms of individual education attainments whose distribution evolve over time with human capital investments. In each period, the distribution of political sophistication within a country generates different incentive structure for the incumbent that accordingly optimally decide whether to be a populist or a responsible type whereas between countries might determine completely different equilibria in the long run, one with populist politicians and one ruled by responsible ones. I argue that rent-maximizer politicians have the chance to behave in a populist fashion when a naive electorate fail in keeping rulers politically accountable. Despite citizens are politically committed to responsible economic policy, naive voters are basically unaware of the politicians intentions providing to the latter opportunities for the manipulation of the economy and the electoral outcome. Populist rulers carry out inefficient investment with the only intent to induce a mean-increasing spread in future distributions of human capital so as to increase electoral consensus based on a naive electorate and to maximize tax revenues based on a few of rich.

Chapter 4

We characterize the statistical law according to which Italian primary school-size distributes. We find that the school-size can be approximated by a log-normal distribution, with a fat lower tail that collects a large number of very small schools. The upper tail of the school-size distribution decreases exponentially and the growth rates are distributed with a Laplace PDF. These distributions are similar to those observed for firms and are consistent with a Bose-Einstein preferential attachment process. The body of the distribution features a bimodal shape suggesting some source of heterogeneity in the school organization that we uncover by an in-depth analysis of the relation between schools-size and city-size. We propose a novel cluster methodology and a new spatial interaction approach among schools which outline the variety of policies implemented in Italy. Different regional policies are also discussed shedding lights on the relation between policy and geographical features.

Chapter 1

Introduction

Almost thirty years have passed by from the time the Nobel prize Robert E. Lucas wrote '*On the Mechanics of Economic Development*'in 1988 in the JME. From that time, several researches have repeatedly stressed the paramount importance of the human capital for the developing of a country. Investment in human capital generate spillover effect on the economy reducing the diminishing return to capital accumulation. According to such literature, differences in growth rates across countries are mainly attributable to differences in the rates at which those countries accumulate human capital over time.

A few has been rather said in Economics about the role of human capital in politics. Bourguignon and Verdier (2000) represents one of the first attempt on this matter. In their model, Oligarchy helps poor in acquiring education to rise up its payoffs via educational externalities. Only educated persons are allowed to take part to political decisions and, as a result, the education process of the poor segment of the society favors the extension of the franchisee that, in turn, lowers the Oligarchy's political power and privileges. In this fashion, education is both the engine of growth and a determinant of political participation.

While the fact that education was a determinant of political participation was a novelty for the economic theory, in Political Science that was a milestone. In 1959 Lipset already identified two mechanisms by which education promotes democracy: (a) education enables a culture of democracy and, at the same time, (b) it leads to greater prosperity, which is also thought to cause political development. What exactly means a *culture of democracy* and the way education directly impacts on democracy is something still debating.

There have been however several empirical works that have documented the existence of a significant (conditional) correlation between education and democracy. Without addressing any causal relation between the two, Barro (1999) and Przeworsky et al. (2000), have provided evidence consistent with the view popularized by Lipset (1959), whereas Glaeser et al. (2004) further investigated the empirical nexus arguing that differences in schooling are a major causal factor explaining not only differences in democracy, but more generally in political institutions. Introducing country fixed effects, Acemoglu et al. (2005) have challenged this view by showing that higher levels of (average) education within a country are not significantly associated with higher achievements in the democratization process.

The point risen by Acemoglu et al. (2005) has been recently reverted by Castelló-Climent (2008) who stresses that what really matters for the implementation and sustainability of democracy is an increase in the education attained by the majority of the population rather than the average years of schooling.

Why does the median of the distribution matter? And why does the mean fail in identifying the relationship between education and democracy within a country? The reason relies on the statistical features of the average, that might suffer from the presence of outliers or because of fat tails in the distribution, and on psychological features involved in the relationship ruler/citizen. In unequal societies, such as Latin American countries, the poorest segment are basically impeded to get educated, because of a liquidity constraint, whereas the Elité attend the best schools in the country optimally investing in human capital. In other countries, where for example the school system is publicly financed, primary and secondary education is mandatory and that does include the poor. In the latter case, (relatively) rich and (relatively) poor attend same schools, contributing to explain why such (more integrated) societies are dynamically better at reducing heterogeneity (see Benabou, 1996; Glomm and Ravikumar, 1992). While the mean hides such differences, the resulting political equilibrium might be surprisingly different among such countries.

The second reason that might contribute to explain why the median matters stands on cognitive process that involve political decisions of every citizens. Education is a cognitive tool that citizens/voters can use to decode the information content of political signal and to keep rulers in charge accountable. While education generates spillover on the economic system, increasing the marginal return to human capital accumulation, it is also deputed to generate political externalities that might favor a culture of democracy. A well-functioning democracy consists of free elections and, once the ruler is in office, of check and balances necessary to maintain that culture. One of the informal check and balances is the political accountability effort that each citizen exerts to control the task of the government. When the majority of the voters is able to control the actions of the rulers, politicians are said to be *accountable* and the accountability turns to be *politically sophisticated*.

Despite political sophistication has been generally ignored as a source of heterogeneity in voting in the Political Economy literature, a number of researchers, within the field of Political Psychology, broadly discussed in Section 2.1, have repeatedly stressed the importance of cognitive sophistication in shaping an individual's ability to make political and economic evaluation (Abramowitz, Lanoue, and Ramesh, 1988; Sniderman, Brody, and Tetlock, 1991; Zallen, 1992; McGraw, 2000; Gomez and Wilson, 2001; Federico and Sidanius, 2002). These works point out that more educated citizens are less likely to be cheated by politicians.

This dissertation consists of three essays. The first two of them address the theoretical nexus between education and democracy. Chapter 2 provides a framework to analyze the relationship between a selfish politician in office and different citizens/voters, endowed by different level of education, that, according to his own level of education, codify the signal (the announcement on public investment) sent by the ruler *differently*. The rent maximizer ruler draws information rents by announcing the wrong state of the world to credulous citizens by claiming that investments are not viable when they actually are. Chapter 2 provides a rationale for the rise and persistence of dissonant rulers (always signaling the wrong state) in naive societies and congruent rulers (always signaling the right state) in sophisticated societies, describing conditions for multiplicity of the equilibria.

Chapter 3 extends this idea to a symmetric scenario in which, tough investment are not feasible, populist and reckless politicians get to office in naive societies by claiming big change for the better are possible and that they can make them happen. Chapter 3 additionally addresses the theoretical puzzle on the sustainability over time of populist regimes and on why the majority bring support to a populist ruler if, in the long run, he leads them to poverty. Multiple equilibria are characterized, one for naive societies ruled by populist rulers and one for sophisticated societies run by responsible politicians.

Should Government emphasize education then? Motivated by the first two Chapters, in Chapter 4 I statically analyze the way the Italian primary school system works and, particularly what kind of education is provided in different part of the Country, despite school autonomy and school competition have nowadays formally granted. Since the distribution of education *matters* for the economy and a well-functioning democracy also the distribution of the school service must be taken into account by the social planner in order to avoid social stratification and inequality. Primary schools tend to distribute in a complex way according to geographical features of the territory, schooling aged population density, and possible interactions between the two. Despite the school system is financed at a State level, Chapter 4 outlines the persistence of remarkable differences not directly attributable to historical divergences among different macro-area of the country but rather between montane areas and more dynamic regions.

Chapter 2

Sophisticated Electoral Accountability: A Political Psychology Agency Theory

2.1 Introduction

Imagine two hypothetical voters. One is exceedingly well informed about politics, a daily and devout reader of the New York Times, who follows closely the major issues of the day, both national and international. The second, a Daily News fan, is hardly overburdened by the amount of time, or effort, she devotes to public affairs – in fact, looks only at the sports page and cares next to nothing about politics. Is it plausible to suppose that these two voters, asked to make a choice about who should be president of the United States, would make up their minds in the same way?

Sniderman, Brody, and Tetlock (1991) at p. 165

Standard political agency models focuses on elections as an incentive devices through which voters discipline politicians. After the seminal papers by Barro (1973) and Ferejohn (1986), second generation models start to combine hidden action and different types of politicians in order to address political selection issues. Politicians may differ among them in their competence (Austen-Smith and Banks, 1989, Banks and Sundaram, 1993), or in their motivation (Besley and Case, 1995, Coate and Morris, 1995, Fearon, 1999, and Rogoff, 1990)[1]. Besley (2006) proposes the distinction between dissonant and congruent rulers, arguing that the latter are more able/willing to give voters what they want.

Although these issues are undoubtedly important, they naturally raise the questions of why a politician should behave congruently or dissonantly and why in some countries rulers perform better than in others. As Figure 2.1 points out, countries around the World widely differ in terms of WGI Government Effectiveness. On one side, developed countries rank on the first quintile of the worldwide distribution. The first ranked country in 2010 is Finland, followed by Singapore, Denmark and Sweden where public and civil services, the degree of its independence from political pressures, the policy formulation and implementation are thought of to be of the highest quality, whereas the government's commitment to such policies the most credible (Kaufmann, Kray, and Mastruzzi, 2010). On the other side, the last ranked countries are mostly developing countries to a great extent located in the sub-Saharan African area (drawn in white in Figure 2.1). What is deputed to explain such a worldwide pattern?

Adding to the puzzle is the fact that most of developing countries leaders are not less competent than those of developed countries, at least in terms of education background, since most of them graduated in U.S. or European universities. This is the case for example of the longest serving ruler of the African continent, Teodoro Obiang Nguema Mbasogo, ruling over the Equatorial Guinea since the 1979 when he overthrew his uncle in a bloody coup d'etat after having successfully completed his studies in Spain. He is followed, in terms of ruling duration, by the President of Angola, José Eduardo dos Santos who brilliantly graduated in engineering in Soviet Union. *Forbes* also includes the King of Swaziland,

[1] For a complete review of the literature see Persson and Tabellini (2000), Grossman and Helpman (2001), Besley (2006), and Ashworth (2012).

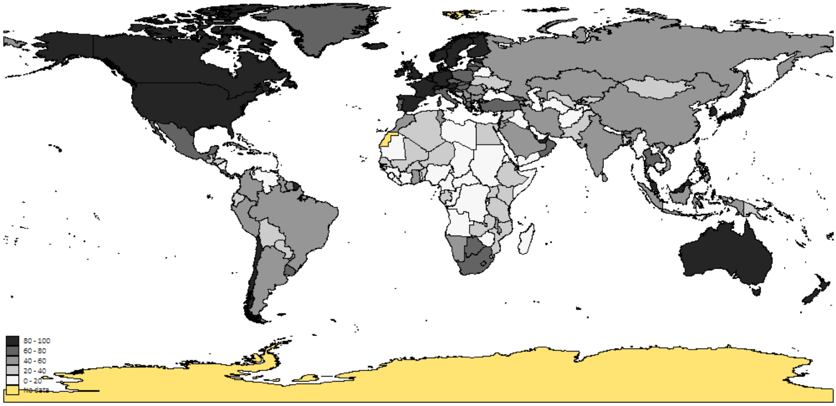


Figure 2.1: The WGI Government Effectiveness worldwide distribution.

King Mswati III, graduated in United Kingdom[2], whereas David Wallechinsky in 'Tyrants: The Worlds 20 Worst Living Dictators'put firmly in discussions the credibility of the elections organized by Robert Biya, President of Cameroon, who years before studied at Sorbonne and Sciences Po in Paris. But many other similar examples involve similar countries across the developing world.

In this paper I argue that rent maximizer politicians rule congruently public business when a sophisticated electorate account for it and behave dissonantly when they are allowed to. Despite political sophistication has been generally ignored as a source of heterogeneity in voting in the Political Economy literature, a number of researchers have repeatedly stressed the importance of cognitive sophistication in shaping an individual's ability to make political and economic evaluation (Abramowitz, Lanoue, and Ramesh, 1988; Sniderman, Brody, and Tetlock, 1991; Zallen, 1992; McGraw, 2000; Gomez and Wilson, 2001; Federico and Sidanius, 2002). These works, related to the Political Psychology literature, point out that more educated citizens are less likely to be cheated by politicians. Education provides political sophistication which is meant to give

[2]From the article 'The Five Worst Leaders In Africa', *Forbes*, September 2, 2012.

individuals ability in making political evaluations. I model Political Psychology predictions by allowing sophisticated voters to perfectly know which is the type of the politician they are facing of (that is whether he is a congruent type or a dissonant one) whatever the signal the latter chooses to send. In this sense, fully sophisticated voters are not involved in any asymmetric information issues that rather interest the rest of the electorate. On the other side, naive voters are basically unaware of the politicians type and intentions providing to the latter opportunities for the manipulation of the economy. However, as far as sophisticated citizens are the majority, manipulation would be hard to be carried out. In between, a continuum of citizens take economic and political decisions driven by their own political sophistication, the prior belief upon the state of nature, and the codification of the signal sent by the politician.

I propose a dynamic signaling political model where citizens/voters are endowed by different level of education and, according to his own level of education, each of them codifies the signal (the announcement on public investment) sent by the ruler *differently* (see Bénabou and Tirole, 2002). In every period, the principal-agent game develops as follow: there are two states of nature about the efficiency level of the State in providing a productive public good. In one of them, a very small amount of the good will be provided either because the State is weak or because exogenous shocks bind fiscal policy[3]. On the other side, productive investments could support private activities, but this possibility is under ruler's full discretion. Once the policymaker starts the office comes to know what is the realization of the state, which is then private information and unknown to citizens. After the information is received

[3]One of the plausible interpretation for that is the theory, proposed by Migdal (1988) and Herbst (2000), and nicely investigated by Acemoglu (2005), according to which weak states have a limited capacity to tax and regulate. On the other end, strong states impose high taxes and invest more. Another plausible interpretation is given by exogenous shocks hitting the economy. Examples are oil supply shocks (that hit western European economies during the seventies), new fiscal rules on budget balance (see the European monetary union), the intervention of international agencies like IMF or World Bank. High public debt levels, inherited from previous governments, may also impede rulers to carry out any public investment. The resulting weakness undermines any discretionary power of the ruler who never has the incentive to cheat the electorate.

he sends a (costless) signal to citizens that, in turn, use the education to screen the plausibility of the announcement. On the basis of the posterior beliefs, citizens optimally choose how much invest in human capital and whether to reelect the incumbent or not. Therefore, we take the screening process, driven by political sophistication and education attainments, to identify the implicit effort that each individual devotes to the accountability of politicians.

Despite no costs are involved in signaling, the cheap talk (or babbling) equilibrium is reached only by sophisticated citizens when the sender behaves dissonantly, always revealing the wrong state. In other words, if the government is looked at as an untrusted one, sophisticated citizens will not pay any attention on the signal it sent. However, in any other cases, the signal still conveys information although the content might be misleading by inducing naive citizens to guess public investment are viable when actually they are not. The ruler does want to cheat citizens, in some state of the world, in order to invest less in a not-directly observable productive public good and appropriating what is left of the tax revenues by rents. However, reducing public investments amounts to shrink tax revenues and, indirectly, his (eventual) future rents heading himself toward a binding trade-off[4]. On the other side, citizens do prefer more public investments that increase individual productivity and, indirectly, the education attainments. The role of citizens as voters is to account for the policymaker job making sure that he invests when he can. Less investments would mean for citizens an income loss, that of course they do want to avoid. Once the incumbent ruler is thought to be dishonest by the majority of the voters he will be punished ex post with another (identical) politician (see also Ferejohn, 1986). Therefore, how democracy works in equilibrium is determined by the accountability effort pushed by voters that is in turn based on the overall level of sophistication of the society.

Unlike the existing works in political agency, I take one step further by endogeneizing agent's types (dissonant versus congruent) and polit-

[4]Something similar is found in Bourguignon and Verdier (2000) where oligarchy helps poor in acquiring education to rise up its payoffs via educational externalities but, at the

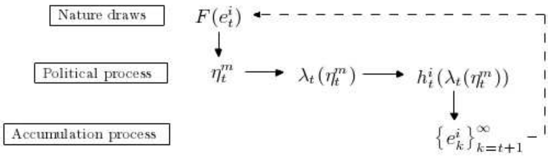


Figure 2.2: The causal relation timing. In every time period, the politician in charge optimally responds to the educational level of the median citizen. e^i_t is inherited from previous periods according to the law of motion $e^i_{t+1} = (1 - delta)e^i_t + h^i_t$. The private investment level in human capital is, in turn, affected by the behavior of the ruler shaping the educational path $\{e^i_k\}^{\infty}_{k=t+1}$ of the society.

ical choices as best responses of the sophistication rate of the electorate. In equilibrium I found the ruler's congruence rate to be increasingly dependent on the educational level of the median voter who is found to be pivotal. This allows us to address the puzzle stated at the beginning explaining away why in some countries rulers perform better than in others. The idea, sketched in Figure 2.2, is that more educated societies are more able to punish politicians that, in turn, invest more in infrastructure, roads or legal rules for contracts enforcement. These productive public goods foster private investment in education (or human capital) making future accountability more effective. The combination of the accumulation and political mechanism creates the potential for multiple steady states, one for low-education societies with dissonant rulers and one for high-education societies with congruent rulers. Similarly to Bénabou (2000), we show how exogenous shocks can move countries from one equilibrium to another. For example, a financial crisis can make the fiscal budget constraint tighter so as to impede new productive public investments. Anticipating that, citizens reduce human capital investments. If the crisis is persistent enough, the resulting lack of investment will lead the society to loose sophistication (the median citizen moves left), giving politicians more chances to behave dissonantly. This mechanism could contribute to explain several historical events such as the advent of totalitarian regimes in the aftermath of the World War I as

same time, extending the franchisee lowers its political power and privileges.

the penalties imposed to Germany by the treaty of Versailles (what John Maynard Keynes defined a Carthaginian peace) can be seen as a huge (persistent enough) fiscal shock.

Behind this mechanism, there is the assumption that more educated citizens are less likely to be cheated by politicians. This is quite in tune with a large strand of literature that crosses behavioral economics, political science, and psychology that stresses the importance of cognitive sophistication in shaping an individual's ability to make political and economic evaluation[5]. Despite political sophistication has been generally ignored as a source of heterogeneity in voting, in much of the political psychology literature, sophistication, awareness, and education are used almost interchangeably (see Sniderman, Brody, and Tetlock (1991) and Zallen (1992)). In this fashion, the ability to keep politicians responsible can be thought of as a function of education, since educated voters are better able to perceive the influence of government policies and macroeconomic fluctuations on their own economic fortunes (Abramowitz, Lanoue, and Ramesh, 1988).

Our main contribution to the existing literature is to provide a general framework that incorporates into the political agency models some aspects of political psychology. Since public investment are not directly observed, both economic and voting decisions are indeed driven by beliefs that are updated through an awareness-management model á la Bénabou and Tirole (1999, 2002). According to that citizens are not *equally* aware about what is going on and only some of them are fully bayesians. We call them sophisticated citizens; sophisticated citizens know what is the congruence rate of the politician, though they can be minoritarian in the society. Asymmetric information about the extent to which rulers are dissonant only involves economic and political decisions of naive citizens, that, when majoritarian, give rent maximizer politicians the chance to

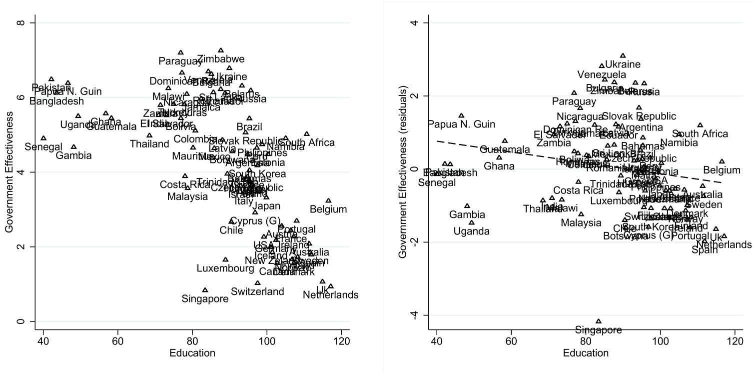
[5]See, among others, Abramowitz, Lanoue, and Ramesh (1988) and more recently Gomez and Wilson (2001). For a complete review on political psychology see McGraw (2000). More related to the political science literature and political business cycle is the pioneer work by Chappel and Keech (1984), that firstly introduces the distinction between naive and sophisticated voters where the former are basically unaware of economic constraints faced by politicians who, in turn, have the opportunities for manipulation of the economy.

manipulate the economy. The model is also general enough to allow several extensions. One of them consider the incentive structure of populist politicians to promise public investments, for electoral purposes, even though there is no room for investing. We also look at the term limit effect in naive societies, arguing that a longer carrier horizon might not be sufficient to induce rulers to behave congruently.

The second contribution relies on the introduction into political agency models of heterogeneity on the principal side. Though heterogeneity complicates things a bit, the model leads to tractable analytic results. In particular, in line with Glomm and Ravikumar (1992) and Bénabou (1996, 2000, 2002), we allow citizens to have different education endowments lognormally distributed. Also the distribution of wealth remains lognormal and closed-form solutions are obtained. Contrary to what has been proposed before in political agency literature, no representative agent has been characterized but, in equilibrium, the median citizen is found to be pivotal, consistently with the median voter theorem. The median citizen also provides a measure of the general level of sophistication (or naivité) of the society leading to a straight intuition of political results and comparative statics.

Our results are partially consistent with the modernization theory that emphasizes the role of education in promoting democracy[6]. On the one hand, education is found to be crucial in shaping democratic institutions via accountability. On the other hand, however, initially low educated societies fail in providing democratic institutions, and, even worst, bad governments are found to be persistent due to a persistent low level of accountability. This endogenous nexus – theoretically developed in this paper – is captured by Figure 2.3, that scatters countries' educational level over the WGI Government Effectiveness index, as measure of good government, for 80 democracies. In Panel a the cross-country unconditional correlation is shown outlining several clusters of countries: consolidated democracies with high levels of education, in the South-East

[6]Lipset (1959) identifies two mechanisms by which education promotes democracy: (a) education enables a culture of democracy and, at the same time, (b) it leads to greater



(a) *WGI Government Effectiveness index.*

(b) *WGI Government Effectiveness index conditional to the GDP per capital level.*

Figure 2.3: Education attainments and government effectiveness for $N = 80$ democracies over the period 1990-98 (observations are averaged over the considered period). Data are taken from Persson and Tabellini (2000). Education attainments are defined as the total enrollment in primary and secondary education, measured as a percentage of the relevant age group in the population, computed dividing the number of pupils (or students) enrolled in a given level of education regardless of age by the population of the age-group which officially corresponds to the given level of education, and multiplying the result by 100. The governance indicators are from Kaufmann et al. (2010). It ranges from around 0 to around 10 (lower values correspond to better outcome). In Panel a I scatter the education attainments over the Government Effectiveness index. In Panel b the Government Effectiveness index is conditioned to the logarithm of the real GDP per capita in 1960. The relationship remains statistically significant.

corner, that more than a century ago have embarked a joint virtuous evolution of institution and political sophistication; minimalist democracies with low level of education, in the North-West corner, that started the democratization process accompanied by a low sophistication ending up in the worst equilibrium (Bidner, Francois, and Trebbi, 2014); countries in between that are still in the middle of the democratization process tending to either equilibria. In Panel b I show that the relationship also stands after controlling for (the logarithm of) the GDP per capita in 1960.

prosperity, which is also thought to cause political development.

There is, in addiction, a wide selection of empirical works that have documented the link between the distribution of education and democracy. Some of them, notably Barro (1999) and Przeworsky et al. (2000), have provided evidence consistent with the view popularized by Lipset (1959), whereas Glaeser et al. (2004) further investigated the empirical nexus arguing that differences in schooling are a major causal factor explaining not only differences in democracy, but more generally in political institutions. Introducing country fixed effects, Acemoglu et al. (2005) have challenged the view that high educational standard is a prerequisite for a country to become a democracy. This conclusion has been recently reverted by two subsequent papers, Bobba and Coviello (2007) and Castelló-Climent (2008). In particular, Castelló-Climent stresses that what really matters for the implementation and sustainability of democracy is an increase in the education attained by the majority of the population rather than the average years of schooling. A measure of the distribution of education has been included in the regression making sure that education attainments are yielded by the less educated fraction of population. This last work is much closer to what we do in this work, given that the median of the distribution defines the general level of sophistication of the society whereas in equilibrium it is found to be pivotal.

The paper is also related to other strands of literature. Beside the aforementioned literature on political agency, there is a growing literature on signaling in elections that draws attention to the role of the politician's platform choice to signal to voters his type (Banks (1990), Harrington (1993)). Kartik and McAfee (2007) study a Hotelling-Downs model of electoral competition where a fraction of candidates have character and are exogenously committed to a campaign platform (different from the median voter position). Callender and Wilkie (2007) develop a general electoral framework in which the willingness to lie varies across candidates and discuss the implications of cheap-talking on signaling equilibria. More recently, Acemoglu, Egorov and Sonin (2013) expand this idea arguing that honest politicians, in order to get reelection, choose populist policies (defined as policies to the left of the median voter) as a way of signaling that he is not beholden to the interest of the rich elite. None of these papers discuss or derive the politicians' credibility or attitude to lie as a best response of the electorate sophistication.

Secondly, the paper relates to the literature that emphasizes the complementarity of the investment by the State and the investment by the citizens. Contributions include Barro (1990), Barro and Sala-i-Martin (1992), Benhabib et al. (2001). We use that framework to shed lights on the importance, in democracy, of good politicians in driving economy through public investment. Bad politicians, on the other side, draw rents wiping out private incentive to invest. The endogenous growth mechanism and the accumulation of human capital has been widely studied by Lucas (1988), Galor and Zeira (1993), Durlauf (1996), Gradstein and Justman (1997), Saint-Paul (1994). Most directly related are the models in Bénabou (1996, 2000, 2002) where producers have different level of human (or physical) capital, lognormally distributed; the accumulation and the redistributive mechanisms dynamically interact pushing different unequal societies to different equilibria (social contracts). Here we build on Bénabou's framework to clarify dynamic interaction between election (or political participation) and the accumulation process. In this sense, our work is also close in spirit to Bourguignon and Verdier (2000) in which an oligarquic society is split up into an initially uneducated poor class that do not participate to political decisions that are only taken by a rich elite; more equal societies democratize sooner because the higher are the incentive for an educated elite to subsidize the poor's education, that, in turn, gain political control. Similarly to Bourguignon and Verdier (2000), we allow education to be both the engine of growth and a determinant of political participation. By doing that we show how more educated societies are more able to punish politicians that, in turn, invest more in productive public goods. In the stationary state, countries initially educated reach an upper bound in education and wealth. Others persist in a ignorance trap[7].

[7] Ashworth, Bueno de Mesquita and Friedenberg (2013) similarly speak of accountability traps that are driven by bad expectations. In an accountability trap a polity is caught in a self-reinforcing pattern of behavior with low accountability and – without changing institutions – another self-reinforcing pattern of behavior with greater accountability and higher voter welfare exists.

The chapter is organized as follows. In Section 2.2 we shall introduce the main features of the model, namely preferences and beliefs, and voting rules. The Markov Perfect Bayesian Equilibrium will be characterized in Section 2.3, firstly discussing how the accumulation process is affected by the institutional setting (the ruler's congruence rate) and then endogeneizing political choices as best responses of the sophistication rate of the electorate (via accountability). We then discuss the effect of inequality on the players' payoffs. Section 2.4 discusses the dynamic implication of the political economic interaction and multiple equilibria are characterized. Section 2.6 concludes.

2.2 The model

2.2.1 Technology and preferences

An incumbent policymaker competes against the opponent to stay in office. In absence of term limits, office guarantees to incumbents a lifetime flow of rents

$$U_0^r = T_0 - BG_0 + \mathbb{E}_0 \sum_{t=1}^{\infty} \beta^t (T_t - BG_t) \varphi_{t-1} \tag{2.1}$$

conditionally of being in office in time t. φ_t is a state variable of the economy, given at time $t = 0$ and equals to 1 if r is the incumbent. Yet, φ_t endogenously evolve over time: at the end of every period t election are held and voters are called for to retain the incumbent or to replace him with a challenger, that is to choose an action $\varphi_t = \{0, 1\}$. Rents are composed by tax revenues T_t the citizens pay to benefit from a productive public good that costs B to the administration, plus future discounted incomes, conditionally of being reelected, i.e. $\varphi_{t-1} = 1$. $G_t = \{0, 1\}$ is an indicator function equals to 1 if the public investment has been made in time t.

The opponent running against the incumbent is identical in all respects from the viewpoint of the voters. Thus the only reason for not reappointing the incumbent is to punish him ex post by taking off future rents, and since the opponent is identical it is indeed (weakly) optimal for the voters to carry out this punishment[8].

The politician in office rules over a continuum of unit mass of risk neutral infinitely-lived citizens endowed with different initial educational levels, e_0^i, normalized to $0 \leq \eta_0^i \equiv e_0^i/\bar{e} \leq 1$, where $\bar{e}$ can be larger or equal to the maximum of the education distribution. The distribution of η_0^i is initially exogenous according to $F(\eta_0^i)$ and evolves (endogenously) across periods on the basis of the following law of motions:

$$e_{t+1}^i = (1-\delta)e_t^i + h_t^i \tag{2.2}$$

where h_t^i is investment (effort) in human capital carried out to accumulate human capital that in turns persists over time with rate $1-\delta$. Despite educational differences, all citizens use the same (Cobb-Douglas) technology to transform individual effort in the unique final good in the economy to be consumed according to (2.4):

$$y_t^i = G_{t-1}^{\gamma}(e_t^i)^{\alpha} \tag{2.3}$$

$$c_t^i \leq (1-\tau_t)y_t^i - \frac{1}{\phi}(h_t^i)^{\phi} \tag{2.4}$$

where $(1/\phi)(h_t^i)^{\phi}$ are convex costs in investing h_t^i in education, given that $\phi \geq 2$. In order to get analytical results, we allow investing to be equally costly for every skilled citizens despite the most skilled ones are expected to exert less effort in learning. Yet, $\gamma, \alpha > 0$ are the elasticity of the public and private investment on the output, respectively[9]: in this fashion, investment by the state is complementary to the investments of citizens (accordingly to Barro (1990), Barro and Sala-i-Martin (1992), Benhabib et al. (2001)). Production is carried out by the citizens, but it depends

[8] This basically amounts to rule out the case in which the ruler provides the public good without being reelected. Furthermore, since challengers are of the *same* type imposing that, when indifferent, voters reelect the incumbent is costless.

[9] In principle, any values of γ could be allowed albeit for $\gamma \leq 0$ citizens do not care in public sector weakening the attention toward politician announcements (for example, $\gamma = 0$ means that public investments are no longer productive). Furthermore, negative values of γ commute G_t to be a bad that is not so far from what has been observing in

on their investments as well as on the quality of the infrastructure, the strength of the law and order, or yet on the legal rules for contract enforcement. All these factors are determined by the public good investments made by the politician and his decision to not carry out public investments sharply leads citizens to not plan any private investments (see Acemoglu (2005)). A proportional taxation scheme is levied by the ruler to collect tax revenues, $T_t = \tau_t y_t$, aimed to invest in a productive public good (i.e. $G_t = 1$) and to remunerate the politician in charge according to (2.1). The investment made in period t will be productive in the subsequent period $t + 1$ so that the accountability effort carried by citizens for having a properly use of public money falls together with an increase in future consumption flows.

Finally, all citizens discount the future with factor β (the same of the politician) and have the same additive (across states and across time) lifetime utility function

$$U_0^i = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t c_t^i \tag{2.5}$$

that only depends on consumption c_t^i, defined above.

2.2.2 Information and beliefs

There are two state of nature about the efficiency level of the State in providing a productive public good, $\sigma_t = \{H, L\}$. In state L, a very small amount of the good will be provided ($G_t^L \approx 0$). On the other side, productive investments could support private activities in state H, but this possibility is under ruler's full discretion (i.e. $G_t^H = \{0, 1\}$)[10]. Once the policymaker starts the office comes to know what is the realization of σ_t, which is then private information and unknown to citizens who have common prior $P(\sigma_t = H) = q$[11]. After the information is received

many developing countries (e.g. by running wars).

[10]Using the Acemoglu (2005) set-up, in $\sigma_t = L$ the state is weak with a limited capacity to tax and regulate (i.e. $\hat{\tau}^L \approx 0$). In $\sigma_t = H$, conversely, strong states may arise imposing high taxes and invest more ($0 << \hat{\tau}^H < 1$). See footnote 3. That means that all the insights of the model will come up from the state H.

[11]The assumption of common prior is not crucial, and is only done to highlight the effect

the policymaker sends a signal to citizens, $\hat{\sigma}_t = \{H, L\}$, assessing the capability of the State in providing G_t. The ruler can have the incentive to cheat citizens by signaling L when the true state is $\sigma_t = H$. In such a way, he can pocket public savings for himself by larger rents[12]. This reasoning also implies that in state L he never has the incentive to cheat, that amounts to say that $P(\hat{\sigma}_t = L|\sigma_t = L) = 1$.

Due to the asymmetric source of information, the role of citizens as voters is to account for the policymaker announcement making sure that he invests less in state L and more in state H[13]. Preventing to be cheated requires a minimum level of sophistication and awareness about the politicians' purposes that not all of them possess. In what follows, we assume, in line with the political psychology literature (discussed in the introduction), that more sophisticated voters are those more educated in society whereas naive people are the uneducated, even though naivete can be exemplified by other characteristics like inexperience, innocence, simplicity. Then, according to his own level of education e_t^i each citizen codifies the signal sent by the ruler *differently*. The idea is that education helps us to convey all the essential informations to make inference. $\eta_t^i = e_t^i/\bar{e}$ captures this intuition:

- $\eta_t^i = 1 \Leftrightarrow e_t^i = \bar{e}$ stands for an individual i who behaves as a perfect bayesian agent with all the information in hands;

- $\eta_t^i = 0 \Leftrightarrow e_t^i = 0$ stands for an individual i who is fully naive and believes that the ruler tell the truth whatever the state of the

of educational level on the inferential process.

[12]This is the short run gain obtained by choosing $G_t = 0$ so as the cost of public investment, BG_t, is zero. In addiction, there is a more subtle gain which operates in the long run: due to imperfect credit markets, not investing today in G_t impedes the citizens to invest, in turn, in human capital which, for every $\delta > 0$, amounts to move the median voter to the left of the distribution becoming more credulous.

[13]The consequences of temporary information asymmetries are investigated also by Rogoff and Sibert (1988) and Rogoff (1990). They argue that politicians use to manipulate certain not-observable macroeconomic policy variables in the imminence of the vote, to increase the probability of being reelected. Herrington (1993) goes further stressing the role of voters' uncertainty on which policy is best, though no one as far as we know shape the role of voters' naivité.

world[14].

In $\sigma_t = H$, the ruler can either tell the truth to citizens (and investing $G_t^H = 1$) or tell a lie ($G_t^H \approx 0$). Each citizen i knows the congruence rate $0 \leq \lambda_t \leq 1$ of the policymaker according to his own level of education. It follows that only the perfect bayesian citizen (i.e. with $\eta_t^i = 1$) can predict with certainty the true rate of the policymaker; less educated citizens can only know a fraction of his intentions, whereas fully naive agents (i.e. with $\eta_t^i = 0$) believe what the ruler tells whatever the state of the world. The probability that he is cheating (or dissonance rate) is *for citizen i* then equal to:

$$P(\hat{\sigma}_t^i = L|\sigma_t = H) = (1 - \lambda_t)\eta_t^i \tag{2.6}$$

whereas the probability that the government says the truth (or congruency rate) is:

$$P(\hat{\sigma}_t^i = H|\sigma_t = H) = 1 - (1 - \lambda_t)\eta_t^i \tag{2.7}$$

where $\lambda_t = 1$ stands for a politician that always tells the truth, and $\lambda_t = 0$ for one that always tells a lie.

Using an awareness-management model à la Bénabou-Tirole (Bénabou and Tirole, 2002) each citizen i assesses the plausibility of the policymaker's claiming: Is it plausible that the state is inefficient (or weak)? Are the policymaker cheating us investing less in education and increasing office rents? The probability that the policymaker will cheat is for citizen i equal to[15]:

$$\begin{aligned} p_t^i(\lambda_t, \eta_t^i, q) &\equiv P(\sigma_t = H|\hat{\sigma}_t^i = L; \lambda_t, \eta_t^i) \\ &= \frac{(1 - \lambda_t)\eta_t^i q}{1 - q + (1 - \lambda_t)\eta_t^i q} \end{aligned} \tag{2.8}$$

It turns out that $(1 - \lambda_t)\eta_t^i \leq 1$ implies that $p^i \leq q$. The babbling equilibrium where $p^i = q$ is reached iff $\lambda_t = 0$ and $\eta_t^i = 1$: if the government is looked at as a untrusted one, citizens will not pay any attention

[14]See Bénabou and Tirole (2002) and Bénabou (2011) for a discussion about bayesian behavior and naivete and the way to model it in a standard microeconomic model.

[15]A number of posterior distributions are drawn in Appendix for different $F(\eta^i)$, priors, and λ. See Figures 2.11-2.15.

on the signal it sent (see Bénabou and Tirole, 1999, 2002) and classify it as a cheap talkers[16]. However, such a level of awareness can be caught by only the most educated agents in society. If λ_t is still null but $\eta_t^i < 1$, the agent will be led to throw away some degree of awareness (or sophistication) in the inferential process. In this sense, the government has some interests in decreasing η_t^i, to make its moves wider.

2.2.3 Voting

Citizens vote retrospectively according to the evidence they have collected on political announces. Since the government's strategy is realized only after the elections, only individual beliefs are involved in the inferential process. Every citizen i processes all the informations collected and votes again for the incumbent if she has no evidence of the fraud, i.e. iff the evidence E_t in favor of the hypothesis p_t^i is not positive[17]:

$$E(p_t^i) \equiv \log\left(\frac{p_t^i}{1-p_t^i}\right) \leq 0 \tag{2.9}$$

which occurs where $p_t^i \leq 1/2$. It in turn means that if the majority of them has no evidence about the cheating move of the policymaker he will be reelected, contingency that occurs when $P(p_t^i \leq 1/2) = F(1/2) \geq 1/2$. Now, given $0 \leq p_t^m \leq q$ with $F(p_t^m) = 1/2$, we require that

$$F\left(\frac{1}{2}\right) \geq \frac{1}{2} = F(p_t^m) \Longleftrightarrow p_t^m \leq \frac{1}{2} \tag{2.10}$$

by monotonicity of $F(\cdot)$. In other words, it turns out that the policymaker won't be reappointed if the median citizen thinks that he is plausibly cheating them. Therefore, if politician cares about reelection he would be willing to push down p_t^m at least to $1/2$. This is of course easier in a society where people can easily be made fools, i.e. in one with a skewed distribution of η_t^i. We summarize this result in Proposition 1:

[16] See Callender and Wilkie (2007) for a discussion on credible and cheap talkers politicians.

[17] Straightforward computation shows that the logit function, $E_t^i = \log(1-\lambda) + \log\eta^i +$

PROPOSITION 1: *Let* $p_t^m(\lambda_t, \eta_t^m, q) \equiv P(\sigma_t = H|\hat{\sigma}_t^m = L; \lambda_t, \eta_t^m)$.

(i) *If* $p_t^m(\lambda_t, \eta_t^m, q) \leq \frac{1}{2}$ *the optimal strategy in the stage game is to play* $\varphi_t = 1$.

(ii) *If* $p_t^m(\lambda_t, \eta_t^m, q) > \frac{1}{2}$ *the incumbent will not be retained (i.e.* $\varphi_t = 0$*).*

2.2.4 Timing of events

The timing of events within every period is as follows:

T1 Nature draws $\sigma_t = \{H, L\}$, that is private information of the ruler. Each citizen inherits e_t^i from the private investment made at time $t-1$ and benefits from the public investment made by the former government, G_{t-1}.

T2 Politician in office chooses the action λ_t, the congruence rate, and, accordingly, invest in a public good (that will be productive in $t+1$), i.e. chooses $G_t = \{0, 1\}$.

T3 Citizens plan to invest h_t^i in human capital based on their beliefs on the ruler's type.

T4 Elections are held (the median citizen chooses $\varphi_t = \{0, 1\}$ based on posterior beliefs).

T5 Payoffs are given by rents and consumptions to politician and citizens respectively.

2.3 The Markov Perfect Bayesian Equilibrium

2.3.1 The political agency problem and the definition of equilibrium

The model features a typical agency problem where politicians in office maximize private rents, expressed by (2.1). In every period t he must decide whether to appear pleasant to voters and being reelected or not, by

$\log(q/1-q)$, is an increasing function of η^i and q, whereas decreases with λ. Once again, choosing which type of politician being (i.e. an admissible value of λ_t). Behaving congruently, for a given distribution of education, raises the chance of being reelected in the next period, but *nothing* says on political choices that the incumbent will take on in the future. In fact, despite the strong incentive the ruler has, any promises cannot be credible, since, in every period t, the ruler has the chance of disregarding the announcement made on G_t and nevertheless being reelected. This commitment problem impedes politician to build a reputation over time[18] and allows us to solve the dynamic game using the Markov Perfect Bayesian Equilibrium (MPBE) concept. The MPBE is defined as a set of Markovian strategies which only depends on the current payoff-relevant states of the economy, $e_t^i \in \mathbb{R}_+ \cup \{0\}$, $\sigma_t \in \Sigma \equiv \{L, H\}$, $G_{t-1} \in \{0,1\}$, and $\varphi_{t-1} \in \{0,1\}$, and on prior actions within the same date, according to the timing of events in 2.4, denoted by $k_t \in \mathcal{K}$; for every possible history, $l^{t,k} \in L^{t,k}$, of the dynamic game up to time t and stage k of the stage game of time t, such a strategies are best responses to each other. The Markovian strategies are also optimal given beliefs, and beliefs are updated using Bayes' rule, according to (2.8).

More formally, for every $q \in [0,1]$ and for each value of the state variable and each combination of prior moves in the stage game given by $\mathcal{K}$, a Markovian strategy mapping

$$s : [0,1] \times \Sigma \times \{0,1\}^2 \times \mathbb{R}_+ \cup \{0\} \times \mathcal{K} \to [0,1] \times \mathbb{R}_+ \cup \{0\} \times \{0,1\}$$

assigns a value for each of the actions: the congruence rate taken on by the ruler, $\lambda_t \in [0,1]$, the amount of private investment made by each citizen, $h_t^i \in \mathbb{R}_+ \cup \{0\}$, and the decision of reelecting the incumbent, $\varphi_t \in \{0,1\}$. We then proceed to determine the equilibrium within each period by backward induction, given e_t^i, σ_t, G_{t-1}, and φ_{t-1} and the beliefs.

---

more educated people collect more evidence upon the job of politicians.

[18]The impossibility of building up a reputation roots with the seminal work by Barro (1973) and it is a milestone of political agency models with no types differences. According to that campaign promises are meaningless, given that lying is costless, and policies are determined only once a candidate is installed in office.

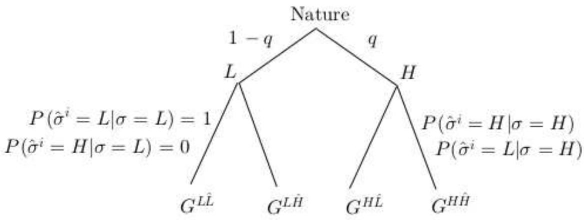


Figure 2.4: Public investments subgame. Note that when the state is thought to be weak every citizens know that the ruler will play $G_t = G^{L\hat{L}} = 0$, no matter the educational level of him. On the contrary, when the state is thought to be strong public investments can be positive ($G^{H\hat{H}} = 1$) or null ($G^{H\hat{L}} = 0$) with probabilities given respectively by equations (2.6) and (2.7).

2.3.2 Elections and political accountability

At the end of each period t, elections are held based on posterior beliefs that each citizen i has. Elections predict that the incumbent will be reelected if the median of the education distribution guesses that he is not cheating them. Reexpressing (2.10) we get the set of admissible strategies for a policymaker who cares about reelection:

$$\lambda_t \geq 1 - \frac{1}{\eta_t^m}\left(\frac{1-q}{q}\right) \equiv \lambda_t^*(\eta_t^m, q) \tag{2.11}$$

According to (2.11), the optimal strategy in the stage game of the ruler is, if the public investment is not too costly, to play in time T2

$$\max\left[0, \lambda_t^*(\eta_t^m, q)\right] \leq \lambda_t \leq 1 \tag{2.12}$$

Note that the fact that $\lambda_t^*(\eta_t^m, q)$ is increasing in η_t^m strongly put into the government's business the issue of accountability. Finally, if $q \leq 1/2$ the inequality (2.11) is always true, for all λ_t (see Figure 2.16).

2.3.3 The accumulation process

We now engage with the accumulation process carried out by citizens given the ruler's strategy λ_t. In T3 citizens invest in human capital on the basis of the informations they have in mind. Accordingly, no private investments would be planned if the state is thought to be weak because $G_t^L \approx 0$ certainly and so will be the output tomorrow. Trivially, each citizen i would carried out $h_t^i = 0$ in any MPBE no matter the education level he has and, in that case, the human capital stock would be firmly the same than period $t-1$ unless depreciation takes place at rate $\delta > 0$. On the other side, with probability q, the state is strong and public investments are thought to be viable; suddenly, each agent's task becomes to puzzle out whether the ruler is telling the truth or a lie (for an exemplification of the cognitive process see Figure 3.1). Using (2.6) and (2.7), we allow politicians to play the following mixed strategy in the stage game:

$$G_t^H = \begin{cases} 1 & \text{with} \quad 1-(1-\lambda_t)\eta_t^i \\ 0 & \text{with} \quad (1-\lambda_t)\eta_t^i \end{cases} \tag{2.13}$$

Thereby the current and the expected future individual output in state H will be

$$y_t^i = G_{t-1}^{\gamma}(e_t^i)^{\alpha} \tag{2.14}$$

$$\mathbb{E}_t(y_{t+1}^i) = q\Big[1-(1-\lambda_t)\eta_t^i\Big](e_{t+1}^i)^{\alpha} \tag{2.15}$$

Recursively, each citizen maximizes the expected current period return that will be consumed according to (2.4) and the agent i's intertemporal utility at time t is

$$\begin{aligned} V(e_t^i) = \max_{h_t^i} \Big\{ (1-\tau)y_t^i - \frac{1}{\phi}(h_t^i)^{\phi} + \beta q\Big[\mathbb{E}_t\big[V^H(e_{t+1}^i)|\hat{\sigma}^i = H\big] + \\ + \mathbb{E}_t\big[V^H(e_{t+1}^i)|\hat{\sigma}^i = L\big]\Big] + \beta(1-q)\Big[\mathbb{E}_t\big[V^L(e_{t+1}^i)|\hat{\sigma}^i = H\big] \\ + \mathbb{E}_t\big[V^L(e_{t+1}^i)|\hat{\sigma}^i = L\big]\Big]\Big\} \end{aligned} \tag{2.16}$$

where income is given by (2.15) and (2.14) and private investments by (2.2). Note that for any $\beta > 0$ each agent cares both in today and tomorrow, and he would like to invest today for consuming tomorrow too. Maximization[19] gives the individual i's optimal investment effort in each

[19] To get analytical results we drop the learning effect by setting investment costs to be

period t as a function of the state variable e_t^i and the ruler congruence rate λ_t:

PROPOSITION 2: *Citizens optimally respond to politicians' congruence rate by lowering private investments when λ_t decreases. The reaction is as strong as larger is the level of sophistication η_t^i:*

$$h_t^i(e_t^i;\lambda_t) = (1-\delta)\left[\frac{\Omega(1-(1-\lambda_t)\eta_t^i)}{1-\Omega(1-(1-\lambda_t)\eta_t^i)}\right]e_t^i \qquad (2.17)$$

with $\Omega(q) \equiv 2\beta q(1-\tau) \leq 1$[20].

Consistently to the endogenous growth literature, Proposition 2 emphasizes the welfare enhancing role of public good provision and thereby the importance to select perfect agents, who always choose $\lambda_t = 1$, to rule public business. As soon as the congruence rate decreases, citizens optimally respond by lowering private investment. However, despite everyone faces the same technology (2.15), the extent of the reaction is heterogeneous. In the appendix (where all the proofs are gathered) we show how $\partial^2 h_t^i/\partial\lambda_t\partial\eta_t^i \geq 0$, meaning that more sophisticated citizens react faster than naives. The former indeed collect all the information required to screen the ruler being aware of what is going on and therefore would be willing to pay more for a given increase in the congruence rate than the naive types.

Education helps them to be informed and sophisticated. However, perhaps surprisingly, we found the effect of education on private investments to be hill-shaped conditional on the politician behavior. Naive citizens always invest increasingly with the human capital stock e_t^i no matter the ruler's strategy unless $e_t^i = 0$ which forces unskilled agents to not invest due to a liquidity constraint for missing credit markets. As soon as more information is acquired, citizens start to be aware of rulers' moves dropping investments if something wrong is thought to

equal to $(h_t^i)^\phi/\phi$. Furthermore, to simplify notations we will keep all the results as functions of e_t^i, dropping G_{t-1}, φ_{t-1} and σ_t as arguments though both are state variables of the economy.

be done[21]. Interestingly, as Figure 2.17 points out, the identified cutoff $\tilde{e}_t(\lambda_t)$ is an increasing function of λ_t; as we showed more formally in the appendix, a congruent political environment wipes out the implications of decreasing educational effect because, for $\lambda_t \to 1$, $\tilde{e}_t$ converges to $\bar{e}$ making investments increasing in human capital for the most skilled agents too. Institutions thus shape the agents' attitudes of investing but the most naive's. The latter always invest more and more even though a dishonest politician has been facing. In this sense the investment reaction of the naive citizens is inelastic with respect to politics: $\tilde{e}_t(0)$ is still positive so as investments are increasing for $e_t^i < \tilde{e}_t(0)$. This is consistent with a large literature dealing with political *extractive* institutions and economic incentives in developing countries (e.g Acemoglu et al., 2001, 2002, 2003; Knack and Keefer, 1995; Hall and Jones, 1999), though heterogeneity and naiveté have never been investigated before.

Besides, complete comparative statics on the subgame equilibrium has been investigated:

(a) First of all, depreciation δ discourages agents to invest more given that much of it will be destroyed in future times ($\partial h_t^i / \partial \delta \leq 0$). Similarly taxation τ does, casting down the accumulation process ($\partial h_t^i / \partial \tau \leq 0$).

(b) On the other side, more optimistic agents clearly invest more ($\partial h_t^i / \partial q \geq 0$) and evenly do the most patient ones ($\beta \to 1$) in order to consume more in the future.

Once the private optimal investment has been characterized, the law of motion of education $e_{t+1}^i(e_t^i; \lambda_t)$, for every citizen i, is easily yielded by substituting (2.17) into (2.2):

$$e_{t+1}^i(e_t^i; \lambda_t) = \frac{1-\delta}{1-\Omega(1-(1-\lambda_t)\eta_t^i)} e_t^i \tag{2.18}$$

[21] Although the absence of learning effects on the accumulation of education is unrealistic, it has the merit to emphasize the drop effect (evenly unrealistically missing in other former models) which encourages more sophisticated citizens to not invest when politicians

State equation (2.18) is an increasing concave function which implies that some mechanisms will lead all the citizens to a common steady state human capital level[22].

2.3.4 The political process

In section 2.3.3 we have demonstrated how the accumulation process and the wealth of a society depends on the institutions and on the extent to which the ruler is willing to be congruent (both summarized by λ). We now move on to consider the reverse interaction going through political mechanisms: how do sophisticated voters bind politician's attitude to be dissonant. This amounts to endogenize political choices as a best response of the sophistication rate of the electorate.

According to the timing of events depicted in section 2.2.4, in T2 the ruler anticipates what is the level of private investments made by each citizen i and chooses the optimal congruence rate $\lambda_t = P(\hat{\sigma} = H|\sigma = H)$ ranged according to (2.12). If he is prone to be congruent to the announcement made in state H he will make the claimed investments. Otherwise, he will not carry any public investment out. More generally, we allow politicians to play the following mixed strategy in the stage game[23]:

$$G_t = \begin{cases} 1 & \text{with} \quad \lambda_t \\ 0 & \text{with} \quad 1 - \lambda_t \end{cases} \tag{2.19}$$

so that the MPBE is obtained by solving, according to equation (2.1), the following recursive optimization problem:

$$\max_{\lambda_t} \quad V_t^r(\lambda_t) = T_t - qB\lambda_t + \beta\mathbb{E}_t\left[V_{t+1}^r(\lambda_t)\right] \quad \text{s.t. (2.12)} \tag{2.20}$$

engage in *per se* rent seeking policies.

[22] As we have argued, this is a straight consequence of the absence of learning effect (see footnote 21). Nevertheless, different common steady state human capital levels would be easily obtained allowing citizens to have different priors. In this fashion, more optimistic citizens (i.e. with q^i higher) would get in equilibrium more.

[23] Note that in equation (2.13) the same mixed strategy has been described from the citizens' point of view whom know what is the level of λ_t according to his own level of sophistication. That makes up the *guessed* probability about whether politician are telling a lie or the truth.

Therefore, political choices shape both current rents, $T_t - qB\lambda_t$, and the income flow that is expected in $t+1$ from reelection. Since credit markets are imperfect, the government cannot spend more than what has been collected by taxing the electorate. It implies that the cost of the project $B \leq T_t$. In what follows, we just express B as a fraction $b \in [0,1]$ of the current tax revenues, i.e. $B = bT_t$, which in turn are equal to $T_t = \tau y_t$, that involve aggregated outcome level $y_t = \int_0^1 y_t^i di$. To keep things easy, in line with Glomm and Ravikumar (1992) and Benabou (1996, 2000, 2002), we suppose education to be initially distributed as a log-normal random variable with mean μ_0 and variance Δ_0^2, i.e. $\ln e_0^i \sim \mathcal{N}(\mu_0, \Delta_0^2)$. However, it is easy to note from (2.18) that the distribution of e_t^i, which is endogenous, remains log-normally distributed over time with mean μ_t and variance Δ_t^2. It follows from (2.14) that also income remains log-normally distributed over time with mean $m_t = 2\mu_t + \Delta_t^2$, i.e. $\ln y_t^i \sim \mathcal{N}(2\mu_t + \Delta_t^2, v_t^2)$.

At the same way, from (2.18) we obtain the difference equation which governs the evolution of the economy, i.e. the law of motion of the aggregate level of human capital:

$$\mu_{t+1} = \mu_t + \Delta_t^2/2 + (\Omega - \delta) - \Omega(1 - \lambda_t)\eta_t \tag{2.21}$$

for small values of $\Omega(1-\lambda_t)\eta_t$ and with $\eta_t \equiv \exp(\mu_t + \Delta_t^2/2)/\bar{e}$. Substituting (2.21) into (2.15) yields the expected output in $t+1$ of the economy:

$$m_{t+1} = m_t + 2(\Omega - \delta) - (1 + 2\Omega)(1 - \lambda_t)\eta_t \tag{2.22}$$

The first two terms describe, respectively, the positive effect of the initial condition of the economy and of exogenous parameters that feature preferences (β), beliefs (q), policies ($-\tau$), and the obsoleteness of the capital stock ($-\delta$). The social cost of cheating, in terms of future income loss, is instead showed in the last term. In particular, equation (2.22) makes clear how dissonant politicians, that always play low values of λ_t, are detrimental to citizens reducing future wealth. Interestingly, the social cost increases with Ω and μ_t by pushing politicians' incentive to draw more rents.

The incumbent's expected rents from being reelected are then equal to:

$$\ln V_t^r(\lambda_t) = \max_{\lambda_t} \Big\{ \ln \tau + 2\mu_t + \Delta_t^2 - qb\lambda_t + \beta \mathbb{E}_t \big[\ln V_{t+1}^r(\lambda_t) \big] \Big\} \quad \text{s.t. (2.12)} \tag{2.23}$$

Maximizing rents amounts to choice an optimal rate of congruence λ_t, ranged according to (2.12). By doing that the incumbent trades off future tax revenues with current rents coming from smaller public investment. Due to the functional form of rents, it is easy to note that $\partial \ln V_t^r / \partial \lambda_t \leq 0$ iff

$$b \geq \frac{\beta(1+2\Omega)\eta_t}{q} \equiv \underline{b}(\beta, q, \tau, \eta_t) \tag{2.24}$$

with $\eta_t \equiv \exp(\mu_t + \Delta_t^2/2)/\bar{e}$. In other words, rents are found to be decreasing with his congruence rate provided that the cost of the public investment, relative to tax revenues, is high enough. Interestingly, the threshold $\underline{b}$ is found to be increasing with β, q, and $-\tau$, meaning that better economic conditions reduce the incumbent's incentives to behave dissonantly. So does a more sophisticated electorate. When condition (2.24) holds, incumbent rulers maximize office rents by pushing down λ_t as much as they can. However, given the accountability effort exerted by voters, the lowest still optimal value is the maximum between $\lambda_t^* = 1 - 1/\eta_t^m((1-q)/q)$ and zero, according to (2.12). We summarize this result in Proposition 3:

PROPOSITION 3: *Political equilibrium. Assume* $\ln e_0^i \sim \mathcal{N}(\mu_0, \Delta_0^2)$. *There exists* $\underline{b}$ *and* $\bar{b}(\eta_t^m)$, *decreasing in* η_t^m, *with* $0 \leq \underline{b} \leq \bar{b}(\eta_t^m) \leq 1$, *such that:*

(i) If $b < \underline{b}$ *the incumbent plays* $\lambda_t = 1$ *and the median voter plays* $\varphi_t = 1$.

(ii) If $b \in [\underline{b}, \bar{b}(\delta)]$ *the incumbent plays* $\lambda_t = \lambda_t^*(\eta_t^m)$ *and the median voter plays* $\varphi_t = 1$.

(iii) If $b > \bar{b}(\delta)$ *the incumbent plays* $\lambda_t = 0$ *and the median voter plays* $\varphi_t = 0$ *(Go-for-Broke).*

When public investments are costly enough it is optimal for rents-maximizing policymakers to set the rate of congruence λ_t to be the lowest possible value. However, due to accountability effort, the optimal congruence rate is:

$$\lambda_t^* = 1 - \frac{1}{\eta_t^m}\Big(\frac{1-q}{q}\Big) \tag{2.25}$$

that is increasing both in q and in the overall society's level of sophistication. In particular, it is worth to note that for every $\eta_t^m \leq (1-q)/q$ the optimal strategy in the stage game for the incumbent is to be fully dissonant (i.e. $\lambda_t^* = 0$).

The cost of the public good drives political decisions for a given distribution of education in the society. When b is very small (i.e. $b < \underline{b}$) there is no incentive to cheat the electorate because investing increases future rents more than what he would have obtained today by choosing $G_t = 0$. However, the cost of the project, while feasible (i.e. $b \leq 1$), could be so high to incentive the ruler to go-for-broke. If $b > \bar{b}$ playing go-for-broke by extracting all the tax revenues strictly dominates $\lambda_t = \lambda_t^*$ (and *a fortiori* any $\lambda_t > \lambda_t^*$). The median voter anticipates that for every $b > \bar{b}$ the incumbent behaves dissonantly (i.e. $\lambda_t < \lambda_t^*$) assigning probability zero on the event that the ruler plays a congruent rate greater than λ_t^*. Consequently, the median voter plays $\varphi_t = 0$ and the ruler, that anticipates this move, goes-for-broke.

In the Appendix we also show that $\bar{b}(\delta)$ is a decreasing function of the depreciation rate of the human capital of the producers so as for high level of δ Go-for-Broke is more likely to be the optimal strategy in the stage game. In particular, there exists a threshold δ^* such that for any $\delta < \delta^*$ Going-for-Broke is not an admissible strategy for the ruler, i.e. $\bar{b}(\delta) > 1$.

In what follows we only consider the most interesting case of $b \in [\underline{b}, \bar{b}(\delta)]$.

2.3.5 The characterization of the MPBE

Once λ_t, and then G_t, are realized citizens come to know what is their pay-off, that is how much they have really invested and what is their production and what will be their human capital level in next periods.

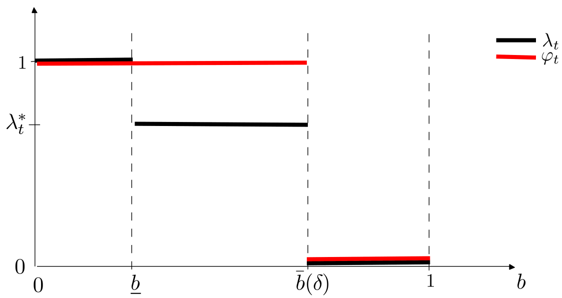


Figure 2.5: Political Equilibrium as a function of b.

This comes out by substituting (2.25) into (2.17) and (2.18) respectively:

$$h_t^i(e_t^i; q, e_t^m) = (1-\delta)\left[\frac{\Omega(1-\frac{e_t^i}{e_t^m}\frac{1-q}{q})}{1-\Omega(1-\frac{e_t^i}{e_t^m}\frac{1-q}{q})}\right]e_t^i \tag{2.26}$$

$$e_{t+1}^i(e_t^i; q, e_t^m) = \frac{1-\delta}{1-\Omega(1-\frac{e_t^i}{e_t^m}\frac{1-q}{q})}e_t^i \tag{2.27}$$

All the MPBE values not only depend on the own level of education but also on the general level of sophistication of the society, e_t^m. This positive external effect works through the accountability effort pushed by the median citizen, that makes more likely a fair political environment with higher investment returns. This is in line with what Bidner and Francois (2013) defines dynamic complementarity between the willingness to vote out today's transgressing leader with a higher expectation that citizens will vote out future transgressors. In our model, dynamic complementarity among voters intertemporal strategies naturally emerge: sophisticated societies force indeed rulers to invest more making them more sophisticated in the future.

Beliefs also play a crucial role, and they are updated according to (2.8); in particular, a strong belief in favor of state L will bring citizens to not carry any investment out. Combining the two effects shows that agents are willing to produce and invest iff $e_t^i \leq (q/1-q)e_t^m$, that is very likely to occur for high values of q (for $q \to 1$ the right hand side diverges to $+\infty$ but a $q = 0$ leads agents to inactivity) and e_t^m.

Therefore we characterize the MBP equilibrium in Proposition 4:

PROPOSITION 4: *Assume* $\ln e_0^i \sim \mathcal{N}(\mu_0, \Delta_0^2)$*, there exists a unique Markov Bayesian Perfect Equilibrium such that*

(i) If the state is weak ($\sigma_t = L$) the incumbent ruler does not carry out any public investment ($G_t = 0$). Citizens know that and respond by not investing ($h_t^i = 0$) regardless of the education they have and by retaining the incumbent ($\varphi_t = 1$); future stocks of human capital are only driven by past level and depreciation (i.e. $e_{t+1}^i = (1-\delta)e_t^i$).

(ii) If the state is strong ($\sigma_t = H$) and if the cost of the public good is sufficiently high, the incumbent ruler has the incentive to cheat citizens that in turn bind politicians by accountability. The ruler cares to be reappointed and the optimal strategy is given by (2.25) when condition (2.24) holds so as citizens relect him ($\varphi_t = 1$). Citizens, on the other side, guess the ruler does not invest ($G_t = 0$) with probability $(1-\lambda_t)\eta_t^i$ and respond by investing and accumulating according to (2.26) and (2.27), respectively.

2.3.6 Equilibrium payoffs and concerns for inequality

State H is the most interesting one, conveying all the insights here presented. Given the optimal strategies, in equilibrium, payoffs are given by future income and rents to citizens and politicians, that in state H are respectively:

$$m_{t+1} = 2\mu_t + \Delta_t^2 + 2(\Omega - \delta) - (1 + 2\Omega)\frac{1-q}{q}\exp\left(\frac{\Delta_t^2}{2}\right) \tag{2.28}$$

$$\begin{aligned} \mathcal{V}_t^r \equiv \ln V_t^r = (1+\beta)\ln\tau + 2(1+\beta)\mu_t + (1+\beta)\Delta_t^2 - qb + b(1-q)\frac{1}{\eta_t^m} \\ + \beta\ln q + 2\beta(\Omega - \delta) - \beta(1+2\Omega)(1-q)\exp\left(\frac{\Delta_t^2}{2}\right) \end{aligned} \tag{2.29}$$

where (2.28) stands for the average producer. Note that both citizens' and ruler's payoffs depend on peculiar features of the distribution of human capital, $\ln(e_t/e_t^m) = \Delta_t^2/2$. In particular, Δ_t^2 describes the extent to which the wealth is unequally distributed among different citizens and, limited to the case of the log-normal distribution, it increases with the mean but declines with the median level of education owned by the political pivotal citizen.

The global effect of inequality on both income and rents turns to be non linear and hill-shaped, meaning that little inequality is tolerated by citizens[24]. The levels of inequality tolerated by citizens, however, are smaller than those preferred by politicians and higher levels hit opposingly citizens and politicians, and only the latter benefit for that. The idea is illustrated in Figure 4, where $\bar{\Delta}^c$ and $\bar{\Delta}^r$ are respectively the bliss points of citizens and rulers, with $\bar{\Delta}^c < \bar{\Delta}^r$. We collect these results in Proposition 5 established in the Appendix:

PROPOSITION 5: *Let* $\ln e_0^i \sim \mathcal{N}(\mu_0, \Delta_0^2)$*, there exists* $\bar{\Delta}^c$ *and* $\bar{\Delta}^r$ *with* $0 < \bar{\Delta}^c < \bar{\Delta}^r$*, such that:*

(i) for each $\Delta_t^2 \in [\bar{\Delta}^c, \bar{\Delta}^r]$ *future income declines with inequality* Δ_t^2 *such that, given the accountability effort, citizens, on average, worst off. Conversely, inequality increases ruler's rents, manipulating poor and extracting rents from taxes of the wealthiest.*

(ii) for $\Delta_t^2 < \bar{\Delta}^c$ *(*$\Delta_t^2 > \bar{\Delta}^r$*) both citizens and politicians better (worst) off with inequality.*

The role of inequality can be easily interpreted into our framework. Positive skewed and unequal (right-tailed) distributions characterize societies with most naive agents and scarcely sophisticated citizens. Naive

[24]This is consistent with the unified theory of inequality and growth proposed by Galor and Moav (2006) who describe a non-linear hill-shaped relationship story. A low-level of inequality enhanced the process of development, mainly driven by physical capital, by channeling resources toward individuals whose marginal propensity to save is higher. In later stages of development, human capital turns to be the main engine of economic growth, and, in the presence of credit constraint, a more equal distribution of income stimulate investment in human capital and promoted economic growth. See also Galor (2011) and Galor and Zeira (1993).

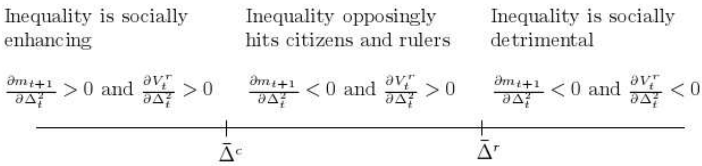


Figure 2.6: Heterogeneous effects of inequality upon future income and politician rents.

agents always invest in human capital and vote for incumbents whatever the latter do. Sophisticated agents, conversely, invest more but under the condition that the incumbent invest too. In this case, they do not reappoint the incumbent either. Because of the positive externalities generated, via accountability, by the median voter, citizens give a positive weight to the median and henceforth to equality, on the region on the right of $\bar{\Delta}^c$, contributing to explain why people dislike living in a society which is too unequal, beside altruism, or aversion to social tension, crime, or civil wars. Pushing down the median value of the distribution instead permits the ruler being more independent from the electorate control. That explains why inequality, on the region on the left of $\bar{\Delta}^r$, is strictly preferred by the rulers. On the other hand, citizens' production is required to bring up tax revenues that, in turns, mainly constitute rents. Too unequal societies, like most developing countries, fail in accumulate human capital that mainly constitutes tax revenues, explaining why rents are hill-shaped with respect to inequality. The interesting social conflict that characterizes well-functioning democracies, with intermediate inequality (i.e. $\Delta_t^2 \in [\bar{\Delta}^c, \bar{\Delta}^r]$), is resolved when inequality is too low or high, and only in the former case it is found to be socially enhancing, since either the median is already high to avoid electorate manipulation or the mean is low, such that more accumulation is wished for.

2.4 Dynamics and multiple steady states

The initial distribution of human capital strongly shapes the dynamics of the economy, political choices, and agents' incentives. Right tailed distributions are mostly composed by naive agents who invest a small but positive amount in human capital and, at the same time, barely account for ruler's duties that, in turn, are allowed to extradraw private rents (by playing small values of λ_t). Conversely, we found that rulers are constrained by more sophisticated societies that impose high values of λ_t as the price for reappointment.

The model thus predicts multiple steady-states, one for sophisticated societies with congruent politicians in charge and one for naive societies ruled by dissonant politicians, and we found the median agent to be pivotal in determining the dynamics of the whole society. To demonstrate that we need to solve the following recursive dynamical system which describes the joint evolution of education and policy:

$$\begin{cases} e^i_{t+1} = \chi(e^i_t; \lambda_t, e^m_t) \\ \lambda_t = \Lambda(e^m_t) \end{cases} \tag{2.30}$$

The solution, at the intersection of the two loci, gives us the long-run human capital level of equilibrium:

$$e_\infty = \left(1 - \frac{\delta}{\Omega}\right)\left(\frac{q}{1-q}\right) e^m_\infty \tag{2.31}$$

where, by equation (2.18), e^m_∞ is a function of the initial condition e^m_0, that is $e^m_\infty(e^m_0)$. The model thus predicts convergence of all agents (but the unskilled ones) to a common educational value, that increases with the median and the prior q. All the parameters that compose Ω, β, q, and $(1-\tau)$, contribute to push up e_∞, whereas it decreases with depreciation. That means that initially less skewed societies, i.e. with higher e^m_0, will be more educated and richer in the long run (see Figure 2.19).

However, as illustrated in Figure 2.19, two societies with the same initial conditions, and in particular with the same initial distribution of education at time $t = 0$, can nonetheless be driven toward two different steady-states. A consistent raise in depreciation or taxation can in fact undermine the effect of optimism (q) and faith in the future (β) persistently discouraging agents to invest to push down society to a zero-level educational state.

The role of the median is then pivotal in that. We show that if $q < \bar{q}(\beta, \delta, \tau)$ the median agent takes a decreasing trajectory and the rest of the society will do the same, firstly the more sophisticated agents and, at last, the naive ones[25]. In fact, if the median decreases over time makes easy the politician to push down λ, that in turn discourage private agents to carry out any investments (see Figure 2.19). More general results are collected in Proposition 6:

PROPOSITION 6: *When citizens are pessimists enough, i.e. if $q < \bar{q}(\beta, \delta, \tau)$, (or equivalently when the depreciation rate is above a critical value $\bar{\bar{\delta}}(\beta, q, \tau)$), society converges to a zero-level of education, no matter initial conditions. Conversely, if $q = \bar{q}(\beta, \delta, \tau)$ multiple stable steady-states arise ($e_\infty = e_0^m$), and rulers are more congruent in societies with higher initial educational achievements. Finally, if $q > \bar{q}(\beta, \delta, \tau)$ society gets richer over time and the speed of the human capital growth is determined by the initial distribution of education.*

Where multiple steady-states occur, *history matters* (Bénabou, 2000). Temporary shocks to the distribution of human capital (such as immigration, educational discrimination, shifts in demand or technology) as well as to the political system (slavery, voting-rights restrictions) can permanently move society from one equilibrium to the other, or more generally have long-lasting effects on the economy.

2.5 Endogenous fiscal choices

In a framework á la Ferejhon (1986) the only instrument that citizens/voters have to punish a bad politician is to vote him out by replacing him with a challenger. Voters decide a political platform or a politician that once in office decide which level of taxation applies. Alternatively, one may think of the tax rate to remain constant over time whoever the politician in power. Redistribution might change only slightly over time whereas

[25]It turns out that $\bar{q}(\beta, \delta, \tau)$ is a U-shaped function of β, bell-shaped in τ and linearly increasing in δ. A society is then less likely to be pushed down to $e_\infty = 0$ for high values of β, small taxes and depreciation.

the willingness to cheat the electorate might change dramatically with a change in office.

We now extend the model to allow agents to decide the optimal fiscal rate, that so far has been held exogenous. Fiscal choices are trivial in state $\sigma_t = L$ and the most preferred rate is $\tau = 0$. Conversely, in state H redistribution is gladly accepted and necessary to provide investment G_t in a productive public good. We then assume that before the ruler announces his decision to carry out any public investment in time T2, citizens are called to decide the most preferred level of redistribution, τ^*. All the other timing is unchanged.

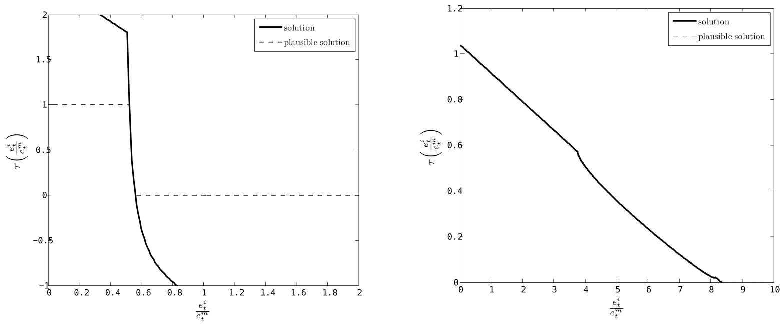
The individual optimal tax rate is obtaining by solving the MPBE by backward induction. Each citizen i, endowed with e_t^i in time t, anticipates which is the optimal congruence rate of the ruler, λ_t^*, and accordingly solve the following quadratic maximization problem, obtained by substituting (2.25), (2.27), and (2.26) into (2.16):

$$\begin{aligned} \mathcal{V}^*(\tau) = \max_{\tau} \Big\{ &(1-\tau)(e_t^i)^2 - \frac{1}{2}(h_t^i)^*(\tau|\lambda_t^*) \\ &+ \beta q [\mathbb{E}_t[V^H((e_{t+1}^i)^*(\tau|\lambda_t^*))|\hat{\sigma}^i = H]] \Big\}. \end{aligned} \tag{2.32}$$

To get analytical results we constrain parameters to have the following values[26]: $q = 0.25$, $\beta = 1$, and $\delta = 0$. Such restriction describes a pessimistic scenario where despite human capital does not depreciate over time and citizens give the greatest importance and weight to future payoffs, the state H is really unlikely to occur. In such scenario, only naive citizens would claim more taxes for doing the public project. The rest of them, and in particular citizens with $e_t^i \geq (1/3)e_t^m$, would like to not contribute to a project that they anticipate will never be done. Figure 2.7(a) shows this scenario.

However, as we pointed out in Section 2.3.4, the optimal congruence rate $\lambda_t^*(q)$ increases with q. Sophisticated citizens anticipate that and, accordingly, choose the optimal tax rate that increases with q. In Figure 2.7(b) the optimal taxation is depicted for $q = 0.75$ as a function of the individual relative level of sophistication. Though naive citizens still

[26] All the algebra is gathered in the Appendix.



(a) $q = 0.25$.

(b) $q = 0.75$.

Figure 2.7: Optimal taxation as a function of political sophistication. We set parameters $\beta = 1$ and $\delta = 0$. In panel (a), a pessimistic scenario is depicted with $q = 1/4$. Here, only naive citizens with $e_t^i \leq (1/3)e_t^m$ would be glad to contribute to a public project that is unlikely to be done. τ^* is not a maximum for $e_t^i \in ((1/3)e_t^m, (9/10)e_t^m)$, but a corner solution has been characterized in the Appendix where these citizens always claim for $\tau^* = 0$. Finally, in panel (*b*) a more optimistic scenario is depicted with $q = 3/4$ and accordingly more sophisticated citizens are more inclined to participate to the public project.

would like to contribute at a higher rate to the public project, sophisticated ones anticipate that the politician in office might be a congruent type and therefore increase τ^* with respect to the former case. Finally, when $q = 1$ all the citizens, regardless the own relative individual level of sophistication, vote for the highest contribution to the project.

We collect this result in Proposition 7:

PROPOSITION 7: *Citizens punish politicians by choosing the optimal contribution rate τ^* according to their own relative individual level of sophistication e_t^i/e_t^m. When they expect politicians to cheat they react by lowering the contribution rate, provided they are sophisticated enough.*

2.6 Conclusions

The paper discusses the importance of education for the success of democracy, as a cognitive tool that citizens/voters can use to decode the information content of political signal and to keep rulers in charge accountable. Remarkably, productivity and citizens welfare may increase with education, via sophisticated electoral accountability, even if education has no direct effect on productivity, via human capital accumulation. This second nexus, tough realistic, is not necessary for the paper story but it only strengthen the results by defining virtuous or wicked political paths, depending on the initial distribution of sophistication in the society.

Aside from the intensity and significance of the political competition and the likelihood of alternation in office, much of the political science literature has pointed out that the most important elements constituting a democracy are the degree of information and participation of the citizens and the structures of accountability characterizing the political sphere. This paper arguably shows that rent maximizer politicians rule congruently public businesses when a sophisticated electorate account for it and behave dissonantly when they are allowed to. Naive voters are basically unaware of the politicians intentions providing to the latter opportunities for the manipulation of the economy. However, as far as sophisticated citizens are the majority, manipulation would be hard to be carried out.

The initial level (condition) of sophistication is striking for the success of a young democracy. More educated societies are more able to punish politicians that, in turn, invest more in infrastructure, roads or legal rules for contracts enforcement. These productive public goods foster private investment in education (or human capital) making future accountability more effective. Such a desirable cycle may fail to start in young democracy when voters are (on a median level) poorly educated, giving rooms to rent-maximizer politicians to cheat them. The combination of the accumulation and political mechanism creates the potential for multiple steady states, one for low-education societies with dissonant rulers and one for high-education societies with congruent rulers.

Promising future works is expected from this preliminary investigation of the behavioral mechanism of political economy. The model is indeed general enough to allow several extensions. One of them consider the incentive structure of populist politicians to promise public investments, for electoral purposes, even though there is no room for investing. On the other side, this framework may contribute to explain why the introduction of term-limits is not fully compatible with rules incentives to pocket public money by investing less. In naive societies, in fact, a longer carrier horizon might not be sufficient to induce rulers to behave congruently.

2.7 Supporting Details

2.7.1 Evolution of beliefs upon the state of the world

Let $\xi = P(\sigma_{t+1}|\sigma_t)$ the persistence rate of the Markov process. From Fig. 2.8 follows that

$$q_{t+1} = P(\sigma_{t+1} = H) = \xi q_t + (1 - \xi)(1 - q_t).$$

The evolution of beliefs upon the state of the world therefore depends on ξ (see Fig. 2.9) for a given level of q_t. One may see that $\xi = 1$ generates a Bernoulli scheme that is a process characterized by fully persistent beliefs, i.e. $P(\sigma_{t+1}) = P(\sigma_t)$. For $\xi = 0.5$ the process is a random walk where future beliefs upon the state of the world are independent from past beliefs but for $\xi = 0$ citizens behave fully irrationally guessing that the state of the world changes time by time. Henceforth, for $\xi < 0.5$ the process is behavioral whereas for $\xi = 0.5$ citizens are totally incapable to make any prediction given q_t[27]. To keep things easy we allow $\xi = 1$ throughout all the paper so as $q_{t+1} = q_t = q$.

[27] That may be a good model for society characterized by high uncertain about future state of the world.

	$\sigma_t = H$	$\sigma_t = L$
$\sigma_{t+1} = H$	ξq_t	$(1-\xi)(1-q_t)$
$\sigma_{t+1} = L$	$(1-\xi)q_t$	$\xi(1-q_t)$

Figure 2.8: Inferential process upon the state of the world .

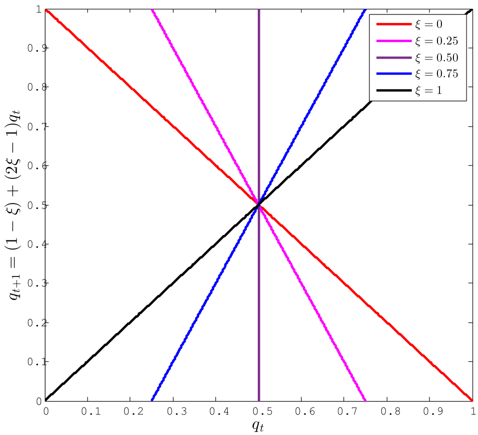


Figure 2.9: Law of motion of $q_{t+1}(q_t)$ for different values of ξ .

2.7.2 Asymptotic features of $e^i_{t+1}(e^i_t, \lambda)$ taking λ as given

Here, we show that the interaction between naiveté and institution acts as a centripetal force with respect to the accumulation process through time leading to a degenerate ergodic distribution in the steady-state. Under several economic conditions and beliefs, the equilibrium is insensitive to political choices and, regardless of sophistication, all citizens get a zero-level of human capital. When q is high enough, conversely, the long run human capital level is positive and increasing in the ruler's congruence rate. We summarize the asymptotic features of (2.18) in Proposition 8, followed by an intuitive proof and discussion. The formal proof is collected in the Appendix.

PROPOSITION 8: *Let* $\Omega(q) \leq 1$.

(i) If q is sufficiently small ($\delta \geq \Omega(q)$), society converges to a zero-level of education ($e_\infty = 0$), regardless of the initial human capital stock and political choices.

(ii) If q is sufficiently high ($\Omega(q) \geq \delta$), society (but the unskilled) converges to the following positive level of education

$$e_\infty(\lambda_\infty) = \frac{\Omega - \delta}{\Omega(1 - \lambda_\infty)}\bar{e} \tag{2.33}$$

that is increasing in λ_∞. The unskilled citizens (with $e_0^i = 0$) are stuck in $e_\infty = 0$ (unstable steady-state) due to a liquidity constraint.

The dynamical interaction between education and institutions underlines an ergodic process where agents can deviate at most temporarily from a common steady-state level of education. λ operates as a centripetal force on the distribution and at last all the skilled agents will have either $e_\infty = 0$, if state L is very likely to occur, or $e_\infty(\lambda_\infty) > 0$, when citizens are optimistic enough. This process has an economic intuition in the latter case. According to his own level of sophistication, each citizen i has an inverse individual measure of dissonance tolerance (IDTM), $\tilde{\lambda}_t^i(\eta_t^i)$ – increasing in η_t^i – such that[28]

$$\partial e_{t+1}^i / \partial e_t^i \geq 1 \qquad \text{iff} \quad \lambda_t \geq \tilde{\lambda}_t^i(\eta_t^i)$$

The statement above tells that only citizens with an IDTM sufficiently small are willing to invest more in the future, that is very likely for naive agents that in turn invest more and more getting more educated (see Figure 2.18). Sophisticated citizens, on the other side, require a congruent political environment to accumulate human capital (λ_t must be high enough). In absence of it they know that much of the investment made will be pocketed by the ruler such that investing more is not optimal till

[28]The extended functional form of $\tilde{\lambda}_t^i(\eta_t^i)$ is gathered in Appendix.

$\tilde{\lambda}^i_{t+j}(\eta^i_{t+j}) \leq \lambda_{t+j}$ for $j > 0$. At last, all the citizens but the unskilled catch up the steady-state $e_\infty(\lambda_\infty)$.

The result collected in Proposition 2 is consistent to the literature that investigates the theoretical nexus between economic growth and development that emphasizes the role of institutions. In particular, this is related to Acemoglu, Johnson and Robinson (2005) that identify economic institutions as a fundamental cause of cross-country differences in prosperity[29]. According to that we introduce Corollary 1 that states that societies with better political institutions, in equilibrium, economically perform better:

COROLLARY 1: *Let $\delta \leq \Omega(q) \leq 1$. Societies with fully congruent rulers ($\lambda_\infty \to 1$) get the possible highest level of wealth $e_\infty \to \bar{e}$. Societies with fully dissonant rulers ($\lambda_\infty = 0$) get the lowest level of wealth $e_\infty = (1 - \delta/\Omega)\bar{e}$.*

2.8 Proofs

2.8.1 Proof of Proposition 2

Claim 1: *there exists a unique optimal level of private investment as a function of the own level of education and the congruence rate of the ruler.*

Proof: The first order condition of maximization (2.16) is

$$(h^i_t)^{\phi-1} = \beta q V'(e^i_{t+1})$$

that equals the costs of investing one unit more today, on the left hand side, to the expected marginal benefits from getting more educated in the future, on the right hand side: these are namely an increasing in future output and the catch up of higher political sophistication. The solution of the FOC uses the standard envelope condition to compute the expected

[29]Interestingly, Acemoglu, Johnson and Robinson (2005) also argue how institutions are endogenous to political power of different social classes that struggle to each other for ruling and having the biggest part of the pie. In the next subsection we endogeneize political institutions through sophisticated accountability.

future marginal benefits:

$$V'_{t+1} = \alpha(1-\tau^H)(1-(1-\lambda_t)\eta_t^i)(e_{t+1}^i)^{\alpha-1}$$

Hence, we are left with

$$(h_t^i)^{\phi-1} = \alpha\beta q(1-\tau^H)(1-(1-\lambda_t)\eta_t^i)(e_{t+1}^i)^{\alpha-1}$$

To get an analytical solution we constrain the human capital elasticity $\alpha = 2$ and investment costs to be quadratic (i.e. $\phi = 2$). This yields equation (2.17). To prove that solution (2.17) is also unique we compute the second order condition that is satisfied iff

$$1 - \Omega(1-(1-\lambda_t)\eta_t^i) \geq 0$$

Since $(1-(1-\lambda_t)\eta_t^i) \leq 1$, a sufficient condition for the SOC to hold is that $\Omega \equiv 2\beta q(1-\tau^H) \leq 1$.

Claim 2: *Citizens optimally respond to politicians' congruence rate by lowering private investments when λ_t decreases.*

Proof: To prove that, we need to demonstrate that $\partial h_t^i/\partial\lambda_t \geq 0$. Let us define $A \equiv \Omega(1-(1-\lambda_t)\eta_t^i)$. Differentiation of (2.17) yields:

$$\begin{aligned}\frac{\partial h_t^i}{\partial\lambda_t} &= \frac{(1-\delta)e_t^i}{1-A}\frac{\partial A}{\partial\lambda_t} + \frac{(1-\delta)Ae_t^i}{(1-A)^2}\frac{\partial A}{\partial\lambda_t} \\ &= \frac{(1-\delta)e_t^i}{1-A}\frac{\partial A}{\partial\lambda_t}\left(1+\frac{A}{1-A}\right) \\ &\geq 0\end{aligned}$$

Claim 3: *The reaction is as strong as larger is the level of sophistication η_t^i (Spence-Mirlees condition).*

Proof: The statement requires that $\partial^2 h_t^i/\partial\lambda_t\partial e_t^i \geq 0$. Straight differentiation yields:

$$\frac{\partial^2 h_t^i}{\partial\lambda_t\partial e_t^i} = \frac{1-\delta}{1-A}\left[2-\Omega\frac{(1-\lambda_t)\eta_t^i}{1-A}\right]\frac{\partial A}{\partial\lambda_t} \geq 0$$

iff $2(1-A) \geq \Omega(1-\lambda_t)\eta_t^i$. Rearranging we get

$$(1-\Omega)+(1-A) \geq 0$$

where both the addends are positive by SOC.

Claim 4: *Private investment are hill-shaped with respect to human capital stock conditional on the political environment.*

Proof: Straight differentiation shows that $\partial h_t^i/\partial e_t^i \geq 0$ iff

$$3\Omega(1-\lambda_t)^2(\eta_t^i)^2 - 2(1-\Omega)(1-\lambda)\eta_t^i + (1-\Omega) \geq 0$$

Solving with respect to e_t^i yields a cut-off value, $\tilde{e}_t(\lambda_t)$, according to that two different patterns of propensity to invest have been found:

- $\forall\, e_t^i \in [0, \tilde{e}_t(\lambda_t)] \Longrightarrow \frac{\partial h_t^i}{\partial e_t^i} \geq 0$
- $\forall\, e_t^i \in [\tilde{e}_t(\lambda_t), \bar{e}] \Longrightarrow \frac{\partial h_t^i}{\partial e_t^i} \leq 0$

Note that the identified cut-off

$$\tilde{e}_t(\lambda_t) = \frac{1}{3}\frac{\bar{e}}{\Omega(1-\lambda_t)}(1-\Omega)^{1/2}\left[(1-\Omega)^{1/2}+(1-4\Omega)^{1/2}\right]$$

is increasing to the congruence rate of the ruler, meaning that a fair political environment wipes out the implications of decreasing educational effect because, for $\lambda_t \to 1$, $\tilde{e}_t$ converges to $\bar{e}$ making investments increasing in human capital for the most skilled agents too. Finally, note that $\tilde{e}_t(0)$ is still positive and increasing for $e_t^i < \tilde{e}_t(0)$, i.e. even with a dissonant politician naive citizens still go through private investment.

2.8.2 Proof of Proposition 3

Assume $\ln e_0^i \sim \mathcal{N}(\mu_0, \Delta_0^2)$. To show that e_t^i keeps distributing lognormally over time let us call

$$\begin{aligned}\xi_0^i &= 1-\Omega(1-(1-\lambda_0)\eta_0^i) \\ &= (1-\Omega)+(\Omega(1-\lambda_0)/\bar{e})e_0^i\end{aligned}$$

Since e_0^i distribute lognormally, with mean μ_0 and variance Δ_0^2, and the other terms in ξ_0^i are constants, ξ_0^i must distribute lognormally as well with mean $\mu_0 + \ln\Omega + \ln(1-\lambda_0) - \ln\bar{e} + (1-\Omega)$ and variance Δ_0^2. Finally, note that the sum of log-normal distributions yields, under certain conditions, another lognormal distribution.

If $\ln e_t^i \sim \mathcal{N}(\mu_t, \Delta_t^2)$ then $y_t^i = (e_t^i)^2$ is a transformation of e_t^i and must distribute as a lognormal too: $\ln y_t^i \sim \mathcal{N}(m_t, v_t^2)$, with $m_t = 2\mu_t + \Delta_t^2$.

The law of motion of aggregated level of human capital is obtained from (2.18). Taking a logarithmic transformation in both sides we are left with

$$\begin{aligned}\ln e_{t+1}^i &= \ln e_t^i + \ln(1-\delta) - \ln(1-\Omega(1-(1-\lambda_t)\eta_t))\\ &\approx \ln e_t^i - \delta + \Omega - \Omega(1-\lambda_t)\eta_t\end{aligned}$$

for small values of δ and $\Omega(1-\lambda_t)\eta_t$. Averaging yields:

$$\mu_{t+1} = \mu_t + \frac{\Delta_t^2}{2} + (\Omega - \delta) - \Omega(1-\lambda_t)\eta_t,$$

with $\eta_t \equiv \exp(\mu_t + \Delta_t^2/2)/\bar{e}$.

At the same way, we can compute the expected output in $t+1$ of the economy; taking a logarithmic transformation of (2.15) yields:

$$\begin{aligned}\ln y_{t+1}^i &= 2\ln e_{t+1}^i + \ln(1-(1-\lambda_t)\eta_t) + \ln(q)\\ &\approx 2\ln e_{t+1}^i - (1-\lambda_t)\eta_t + \ln(q)\end{aligned}$$

Taking the average:

$$\begin{aligned}m_{t+1} &= 2\mu_{t+1} - (1-\lambda_t)\eta_t\\ &= 2\mu_t + \Delta_t^2 + 2(\Omega-\delta) - 2\Omega(1-\lambda_t)\eta_t - (1-\lambda_t)\eta_t\\ &= m_t + 2(\Omega-\delta) - (1+2\Omega)(1-\lambda_t)\eta_t\end{aligned}$$

In time T2 the incumbent ruler faces the following maximization problem:

$$V_t^r(\lambda_t) = \max_{\lambda_t}\left\{\tau y_t - qb\tau y_t\lambda_t + \beta\mathbb{E}_t\left[V_{t+1}^r(\lambda_t)\right]\right\} \quad \text{s.t. (2.12)}$$

Taking a logarithmic transformation yields:

$$\begin{aligned}
\ln V_t^r(\lambda_t) &= \max_{\lambda_t} \Big\{ \ln\tau + \ln y_t + \ln(1-qb\lambda_t) + \beta\mathbb{E}_t\big[\ln V_{t+1}^r(\lambda_t)\big]\Big\} \quad \text{s.t. (2.12)}\\
&\approx \max_{\lambda_t} \Big\{ \ln\tau + \ln y_t - qb\lambda_t + \beta\mathbb{E}_t\big[\ln V_{t+1}^r(\lambda_t)\big]\Big\} \quad \text{s.t. (2.12)}\\
&= \max_{\lambda_t} \Big\{ \ln\tau + 2\mu_t + \Delta_t^2 - qb\lambda_t + \beta\mathbb{E}_t\big[\ln V_{t+1}^r(\lambda_t)\big]\Big\} \quad \text{s.t. (2.12)}
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}_t\big[\ln V_{t+1}^r(\lambda_t)\big] &= \ln q + \ln\tau + m_{t+1}\\
&= \ln q + \ln\tau + 2\mu_t + \Delta_t^2 + 2(\Omega-\delta)\\
&- 2\Omega(1-\lambda_t)\eta_t - (1-\lambda_t)\eta_t
\end{aligned}$$

Since the program is linear in λ_t, it is easy to note that

$$\frac{\partial \ln V_t^r}{\partial\lambda_t} = \beta(1+2\Omega)\eta_t - qb$$

that is negative if and only if condition (2.24) is satisfied, that is if the cost of the public investment, relative to tax revenues, is high enough.

We now show that there exists a cutpoint-cost of the public good $\bar{b}$ such that for every $b > \bar{b}$ the ruler strictly prefer to going-for-broke instead of playing a strategy $\lambda_t \geq \lambda_t^*$. We first show that, at time $T2$, the value function of going-for-broke is strictly greater than the value taken by being congruent enough (we show it for $\lambda_t = \lambda_t^*$; a fortiori it hold for any $\lambda_t > \lambda_t^*$). Let us call $\mathcal{V}_t^r \equiv \ln V_t^r(\lambda_t^*)$.

$$\mathcal{V}_t^r < T_t$$

$$\ln\tau + 2\mu_t + \Delta_t^2 > (1+\beta)\ln\tau + 2(1+\beta)\mu_t + (1+\beta)\Delta_t^2 - qb + b(1-q)\frac{1}{\eta_t^m} +$$
$$+\beta\ln q + 2\beta(\Omega-\delta) - \beta(1+2\Omega)(1-q)\exp\left(\frac{\Delta_t^2}{2}\right)$$

Solving by b yields $\bar{b}$:

$$\bar{b}(\delta) \equiv \frac{\beta\big[\ln\tau q + m_t + 2\beta(\Omega-\delta) - \beta(1+2\Omega)(1-q)\exp\big(\frac{\Delta_t^2}{2}\big)\big]}{q-(1-q)\frac{1}{\eta_t^m}}.$$

Note that $\bar{b}(\delta)$ is a decreasing function of the depreciation rate of the human capital of the producers so as for high level of δ Go-for-Broke is more likely to be the optimal strategy in the stage game. In particular, there exists a threshold δ^* such that for any $\delta < \delta^*$ Going-for-Broke is not an admissible strategy for the ruler. To show that we need to solve the following inequality:

$$\bar{b}(\delta) > 1,$$

that holds for

$$\delta < \Omega - \frac{1}{2\beta}\left[q - (1-q)\frac{1}{\eta_t^m} - \beta \ln \tau q + \beta m_t - \beta(1+2\Omega)(1-q)\exp\left(\frac{\Delta_t^2}{2}\right)\right]$$
$$\equiv \delta^*$$

2.8.3 Proof of Proposition 5

Assume $\ln e_t^i \sim \mathcal{N}(\mu_t, \Delta_t^2)$, such that $\ln(e_t/e_t^m) = \Delta_t^2/2$. Given condition (2.24), the optimal congruence rate is

$$\lambda_t^* = 1 - \frac{1-q}{q}\frac{1}{\eta_t^m}.$$

Substituting it into the average future income m_{t+1} and the ruler's rent V_t^r yields equations (2.28) and (2.29), respectively.

Claim 1: *Income is hill-shaped with respect to inequality.*

Proof: Straight differentiation shows that income is increasing w.r.t. inequality iff:

$$\frac{\partial m_{t+1}}{\partial \Delta_t^2} \geq 0 \Longleftrightarrow 1 \geq \left(\frac{1}{2} + \Omega\right)\left(\frac{1-q}{q}\right)\exp\left(\frac{\Delta_t^2}{2}\right)$$
$$\Longleftrightarrow \Delta_t^2 \leq 2\ln(q) - 2\ln(1-q) + 2\ln(2) - 2\ln(1+2\Omega) \equiv \bar{\Delta}^c$$

Claim 2: *Politicians rents are hill-shaped with respect to inequality.*

Proof:

Straight differentiation shows that politicians rents are increasing w.r.t. inequality iff:

$$\frac{\partial \ln V_t^r}{\partial \Delta_t^2} \geq 0 \Longleftrightarrow \frac{1+\beta}{\beta} \geq \frac{1}{2}(1+2\Omega)\frac{1-q}{q}\exp\left(\frac{\Delta_t^2}{2}\right)$$

$$\Longleftrightarrow \Delta_t^2 \leq 2\ln 2 + 2\ln(1+\beta) - 2\ln\beta + 2\ln q - 2\ln(1-q) - 2\ln(1+2\Omega) \equiv \bar{\Delta}^r$$

Claim 3: *Citizens bliss point is smaller than ruler's.*

Proof: We need to demonstrate that $\bar{\Delta}^c < \bar{\Delta}^r$. It comes out from the definitions in Claim 1 and 2:

$$\bar{\Delta}^c < \bar{\Delta}^r \Longleftrightarrow \ln(1+\beta) > \ln\beta$$

that is always true.

2.8.4 Proof of Proposition 6

A stable steady state is a point $(e_\infty, \lambda_\infty)$ with the curve $\chi(\lambda)$ cuts the curve $\Lambda(e)$ from above. An unstable steady state corresponds in each case to an intersection from below.

The dynamical system (2.27) reduces to a one-dimensional recursion: $e_{t+1}^i = \chi(e_t^i, \Lambda(e_t^i))$. It has the following features:

(i)

$$\chi(0) = 0$$

(ii)

$$\chi'(e_t^i) = \frac{(1-\delta)(1-\Omega)}{\left[1-\Omega\left(1-\frac{e_t^i}{e_t^m}\frac{1-q}{q}\right)\right]^2} \geq 0 \quad \text{by SOC}$$

(iii)

$$\chi''(e_t^i) = -\frac{(1-\delta)(1-\Omega)}{\left[1-\Omega\left(1-\frac{e_t^i}{e_t^m}\frac{1-q}{q}\right)\right]^3}\frac{\Omega}{e_t^m}\frac{1-q}{q} \leq 0 \quad \text{by SOC}$$

(iv)

$$\chi'(e_t^i) \geq 1 \Longleftrightarrow$$

$$\begin{cases} e_t^i \in \left[0, \frac{q}{1-q}\frac{e_t^m}{\Omega}(1-\Omega)^{1/2}\Big((1-\delta)^{1/2}-(1-\Omega)^{1/2}\Big)\right] & \text{iff} \quad \Omega \geq \delta \\ e_t^i \in \left[\frac{q}{1-q}\frac{e_t^m}{\Omega}(1-\Omega)^{1/2}\Big((1-\delta)^{1/2}-(1-\Omega)^{1/2}\Big), 0\right] & \text{elsewhere} \end{cases}$$

(v)

$$e_{t+1}^i = e_t^i = e_\infty$$

in four fixed points:

$$e_\infty^{(1)} = 0$$

$$e_\infty^{(2)} = \begin{cases} \Big(1-\frac{\delta}{\Omega}\Big)\Big(\frac{q}{1-q}\Big)e_\infty^m \geq 0 & \text{iff} \quad \Omega \geq \delta \\ \Big(1-\frac{\delta}{\Omega}\Big)\Big(\frac{q}{1-q}\Big)e_\infty^m \leq 0 & \text{elsewhere} \end{cases}$$

A fixed point e_∞ is stable if and only if

$$\frac{\mathrm{d}\chi(e)}{\mathrm{d}e}\bigg|_{e=e_\infty} < 1$$

Computation tells us that $e_\infty^{(2)}$ is the only stable fixed point if $\Omega(q) \geq \delta$ (or if $q \geq \bar{q}(\beta, \delta, \tau) \equiv \frac{1}{2}\frac{\delta}{\beta(1-\tau)}$), whereas $e_\infty^{(1)} = 0$ is unstable. In the latter case ($\Omega(q) \leq \delta$), $e_\infty^{(1)} = 0$ is the unique stable fixed point.

In the former case, $e_\infty^{(2)}$ depends on the trajectory of the median agent of the distribution, that is then pivotal. In time 1, from (2.27), is evident that $e_1^i \geq e_0^i$ if the citizen i is naive enough, i.e. if

$$e_0^i \leq \Big(1-\frac{\delta}{\Omega}\Big)\Big(\frac{q}{1-q}\Big)e_0^m$$

More sophisticated agents instead will be driven by the political process to decrease the investment in human capital. This reasoning implies a convergence process according to which there exists a catching-up period $t^* > 0$ such that $e_{t^*}^i = e_{t^*}^m$ and $F(e_{t^*}^i)$ is degenerate for every $t \in [t^*, \infty)$. The dynamics of the system turns into a degenerate ergodic process described by the following linear law of motion:

$$e_t = \left[\frac{1-\delta}{1-\Omega\big(1-\frac{1-q}{q}\big)}\right]^t e_0^m \qquad \forall t \in [t^*, \infty)$$

Let us call $D = \frac{1-\delta}{1-\Omega\left(1-\frac{1-q}{q}\right)}$. It follows that e_∞ diverges for every $D > 1$, i.e. for every

$$q > \frac{2\beta(1-\tau)+\delta}{4\beta(1-\tau)} \equiv \tilde{q}(\beta, \delta, \tau)$$

whereas it converges to zero for every $D < 1$. Finally it converges to e_0^m for every $D = 1$. Note that $\tilde{q} \geq \bar{q}$ meaning that $D > 1$ when $\Omega > \delta$.

2.8.5 Proof of Proposition 7

The individual optimal tax rate is obtaining by solving the MPBE by backward induction. Each citizen i, endowed with e_t^i in time t, anticipates which is the optimal congruence rate of the ruler, λ_t^*, and accordingly solve maximization (2.32).

FOC requires $\mathcal{V}'(\tau) = 0$. Since the maximization problem is quite complex we constrain parameters to get analytical results to the following values: $\beta = 1$ and $\delta = 0$. We also initially set $q = 0.25$ to consider a pessimistic scenario, i.e. one in which politicians is rationally expeted to cheat, but later we allow q to vary. We also define $\epsilon^i = e_t^i/e_t^m$ to be the individual relative level of political sophistication.

In this scenario the only plausible solution is given by

$$\tau^*(\epsilon^i|q = 0.25) = 2 + \frac{3}{8}(1 - 3\epsilon^i) +$$
$$+ \frac{1}{2}\frac{\sqrt{\frac{9}{16}(1-3\epsilon^i)^4 + \frac{3}{8}(1-3\epsilon^i)^3 + 17(1-3\epsilon^i)^2 - 20(1-3\epsilon^i) - 16}}{1 - 3\epsilon^i}$$

depicted in Fig. 2.7(a).

SOC tells us that $\tau^*(\epsilon^i|q = 0.25)$ is a maximum for any $\tau^*(\epsilon^i|q = 0.25)$ satisfying the following inequality:

$$\mathcal{V}''(\tau^*(\epsilon^i|q = 0.25)) < 0$$

$$\mathcal{V}'' = (1 - 3\epsilon^i)\Big[\frac{1}{2}(1 - 3\epsilon^i)\big(\tau^* - \frac{3}{8}\big) - 1\Big] < 0.$$

The SOC problem is illustrated in Fig. 2.10. As one may see $\tau^*(\epsilon^i|q = 0.25)$ is not a maximum for $\epsilon^i \in (1/3, 9/10)$ for neither solutions. Then we proceed by computing corner solutions. We limit to show it for $\epsilon^i = 1/2$, but same results apply for any $\epsilon^i \in (1/3, 9/10)$. It is straightforward to see that

$$\mathcal{V}(\tau = 1|\epsilon^i = 0.5, q = 0.25) = 0,$$

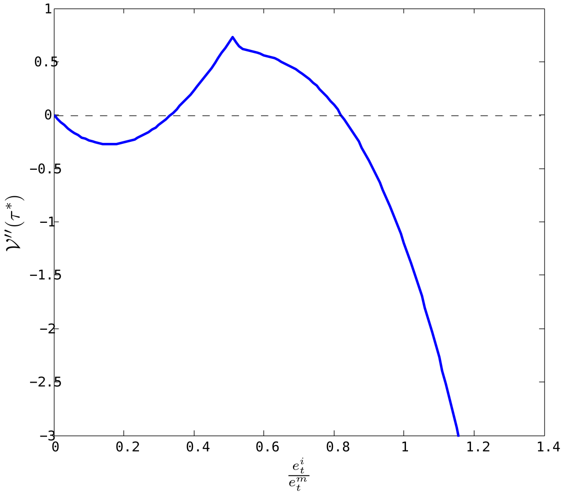


Figure 2.10: Second order condition of the taxation maximization problem.

whereas

$$\mathcal{V}(\tau = 0|\epsilon^i = 0.5, q = 0.25) = \frac{9}{10}(e_t^i)^2 > 0.$$

Since $\mathcal{V}(\tau|\epsilon^i = 0.5, q = 0.25)$ is a monotonically decreasing function and

$$\mathcal{V}(\tau = 0|\epsilon^i = 0.5, q = 0.25) > \mathcal{V}(\tau = 1|\epsilon^i = 0.5, q = 0.25),$$

we conclude that $\tau = 0$ is a corner solution for $\epsilon^i = 0.5$.

A more optimistic scenario, i.e. one in which the ruler is expected to be congruent, is one with $q = 0.75$. In this scenario the only plausible solution is given by

$$\tau^*(\epsilon^i|q = 0.75) = \frac{2}{3} + \frac{3}{8}(1 - \frac{1}{3}\epsilon^i) + \\ + \frac{1}{2}\frac{\sqrt{\frac{9}{16}(1-\frac{1}{3}\epsilon^i)^4 + \frac{9}{2}(1-\frac{1}{3}\epsilon^i)^3 + \frac{25}{9}(1-\frac{1}{3}\epsilon^i)^2 - \frac{20}{3}(1-\frac{1}{3}\epsilon^i) - \frac{16}{9}}}{1-\frac{1}{3}\epsilon^i}$$

depicted in Fig. 2.7(b). SOC also shows that $\mathcal{V}''(\tau^*(\epsilon^i|q = 0.75)) < 0$ stands for any ϵ^i.

Finally one may see that for $q = 1$ each citizen, regardless of his own level of sophistication, knows that the politician in office is a perfect agent always playing $\lambda_t^* = 1$. In such scenario

$$\tau^*(q=1) = \frac{7}{2} \pm \frac{1}{2}\sqrt{29} > 1.$$

2.8.6 Proof of Proposition 8

First of all, we show that $e_{t+1}^i(e_t^i)$ is an increasing concave function. To see that note:

$$\frac{\partial e_{t+1}^i}{\partial e_t^i} = \frac{(1-\delta)(1-\Omega)}{\left[1-\Omega(1-(1-\lambda_t)\eta_t^i)\right]^2} \geq 0$$

$$\frac{\partial^2 e_{t+1}^i}{\partial^2 e_t^i} = -\frac{(1-\delta)(1-\Omega)}{\left[1-\Omega(1-(1-\lambda_t)\eta_t^i)\right]^3}\frac{\Omega(1-\lambda_t)}{\bar{e}} \leq 0$$

where both the inequalities hold by SOC, i.e. $\Omega \leq 1$. Furthermore, it lays above the $45°$-line ($\partial e_{t+1}^i / \partial e_t^i \geq 1$) for

$$e_t^i \leq \left[(1-\delta)^{1/2} - (1-\Omega)^{1/2}\right]\frac{(1-\Omega)^{1/2}}{\Omega(1-\lambda_t)}\bar{e} \equiv \bar{e}_t(\lambda_t)$$

or, equally, for $\lambda_t \geq \tilde{\lambda}_t^i(\eta_t^i)$, defined as

$$\lambda_t \geq 1 - \left[(1-\delta)^{1/2} - (1-\Omega)^{1/2}\right]\frac{(1-\Omega)^{1/2}}{\Omega\eta_t^i} \equiv \tilde{\lambda}_t^i(\eta_t^i)$$

which is an increasing function of η_t^i if and only if $\Omega \geq \delta$. Therefore by concavity we get, substituting in (2.18) $e_{t+1}^i = e_t^i = e_\infty$, two steady-state equilibria. If $\Omega \geq \delta$,

$$\begin{cases} e_\infty = 0 \qquad (unstable) \\ e_\infty = \frac{\Omega-\delta}{\Omega(1-\lambda_\infty)}\bar{e} \geq 0 \qquad (stable) \end{cases}$$

Otherwise,

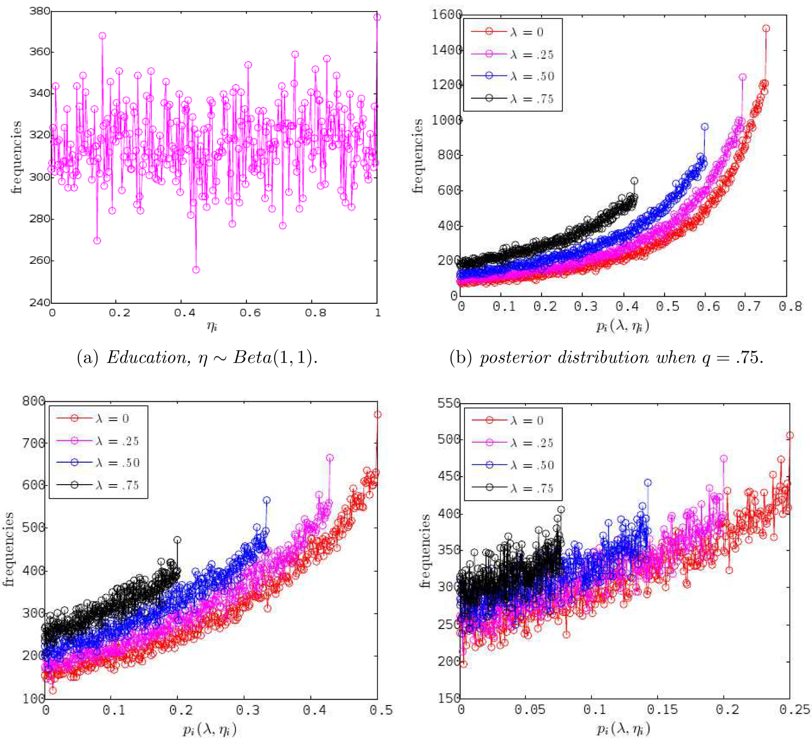
$$\begin{cases} e_\infty = 0 \qquad (stable) \\ e_\infty = \frac{\Omega-\delta}{\Omega(1-\lambda_\infty)}\bar{e} \leq 0 \qquad (unstable) \end{cases}$$

and all the citizens converge to a zero-level of human capital.

Finally, taking the limit of e_∞ with respect to λ_∞ we get, in the most interesting case of $\Omega \geq \delta$:

$$\lim_{\lambda_\infty \to 0} e_\infty = \left(1 - \frac{\delta}{\Omega}\right)\bar{e}$$

$$\lim_{\lambda_\infty \to 1} e_\infty = \bar{e}$$



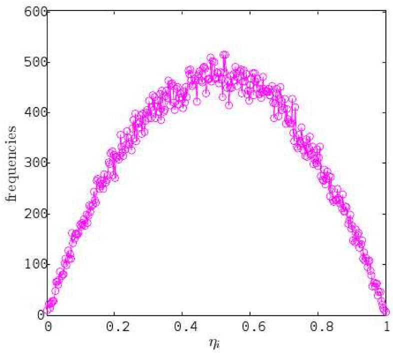
(a) *Education*, $\eta \sim Beta(1,1)$.

(b) *posterior distribution when* $q = .75$.

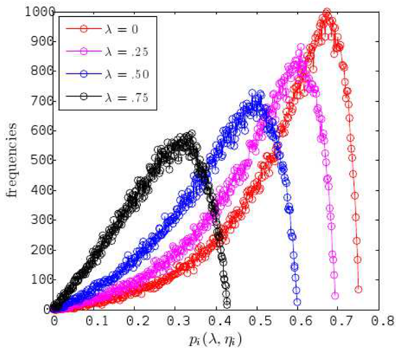
(c) *posterior distribution when* $q = .50$.

(d) *posterior distribution when* $q = .25$.

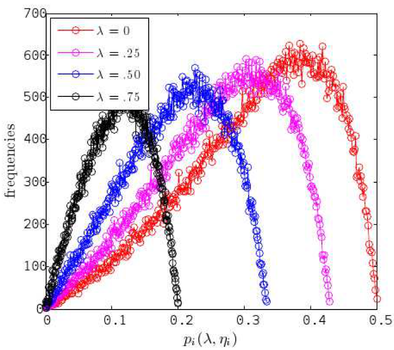
Figure 2.11: Bayesian inference in a society of $n = 10^5$ citizens where education distributes uniformly.



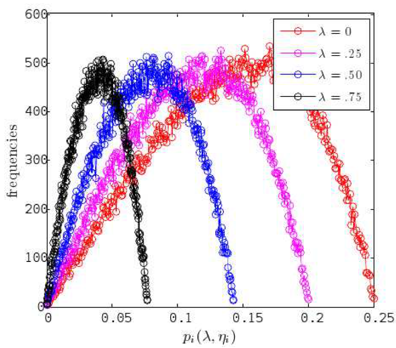
(a) *Education*, $\eta \sim Beta(2, 2)$.



(b) *posterior distribution when* $q = .75$.

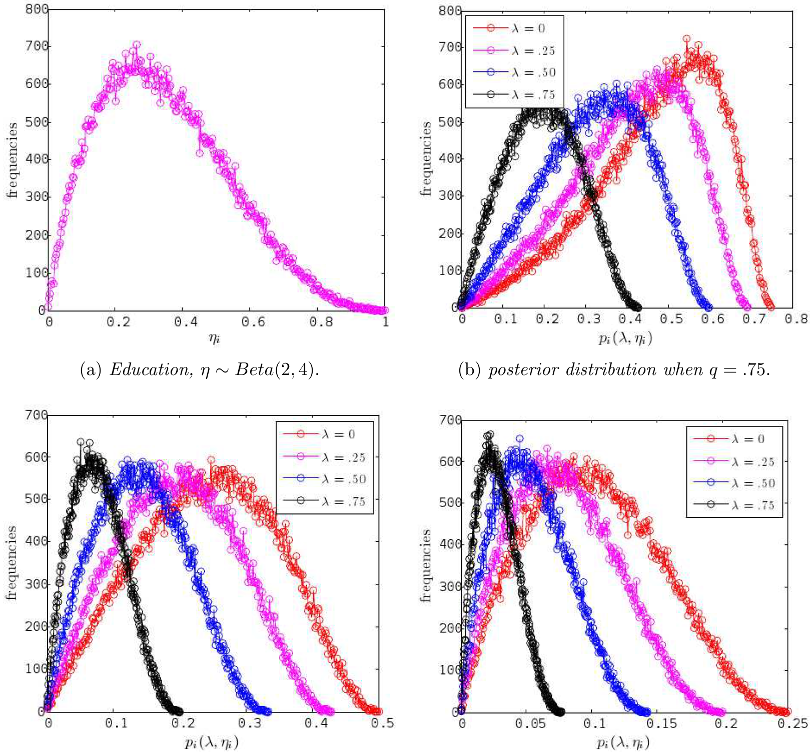


(c) *posterior distribution when* $q = .50$.



(d) *posterior distribution when* $q = .25$.

Figure 2.12: Bayesian inference in a society of $n = 10^5$ citizens where education distributes symmetrically.



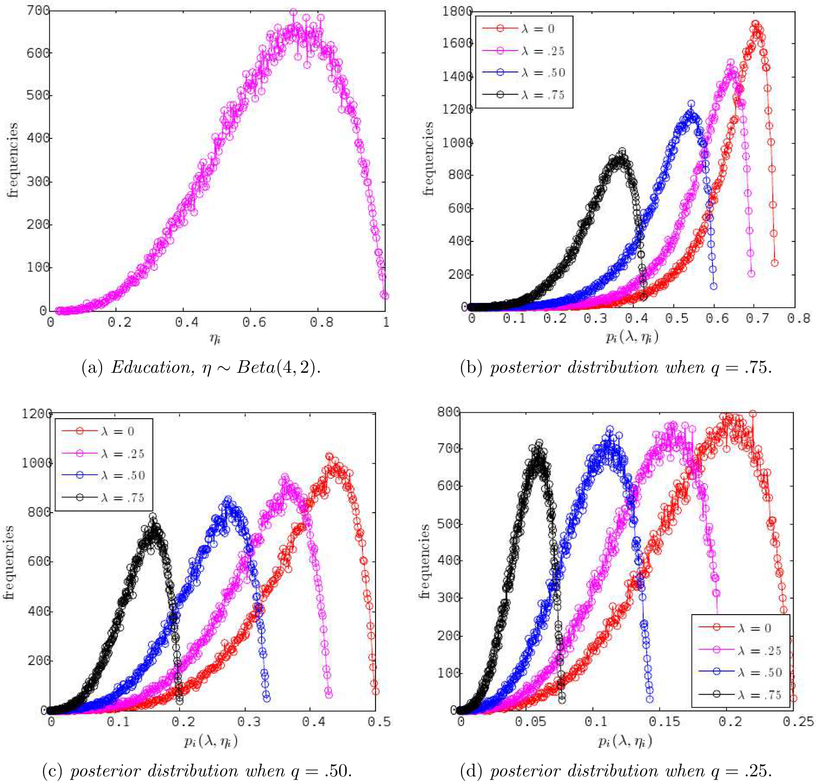
(a) *Education,* $\eta \sim Beta(2,4)$.

(b) *posterior distribution when* $q = .75$.

(c) *posterior distribution when* $q = .50$.

(d) *posterior distribution when* $q = .25$.

Figure 2.13: Bayesian inference in a society of $n = 10^5$ citizens where the distribution of education is positive skewed (right-tailed).



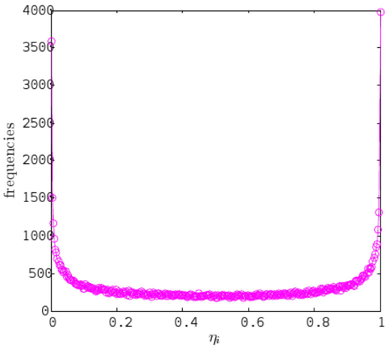
(a) *Education,* $\eta \sim Beta(4, 2)$.

(b) *posterior distribution when* $q = .75$.

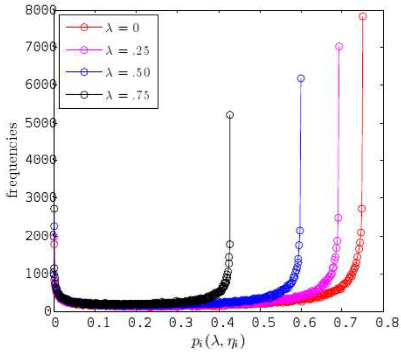
(c) *posterior distribution when* $q = .50$.

(d) *posterior distribution when* $q = .25$.

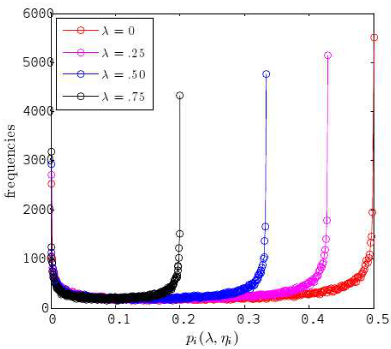
Figure 2.14: Bayesian inference in a society of $n = 10^5$ citizens where the distribution of education is negative skewed (left-tailed).



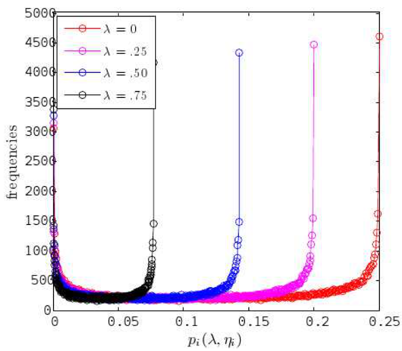
(a) *Education,* $\eta \sim Beta(.5, .5)$.



(b) *posterior distribution when* $q = .75$.



(c) *posterior distribution when* $q = .50$.



(d) *posterior distribution when* $q = .25$.

Figure 2.15: Bayesian inference in an unequal society of $n = 10^5$ citizens where the distribution of education is symmetric but fat tailed.

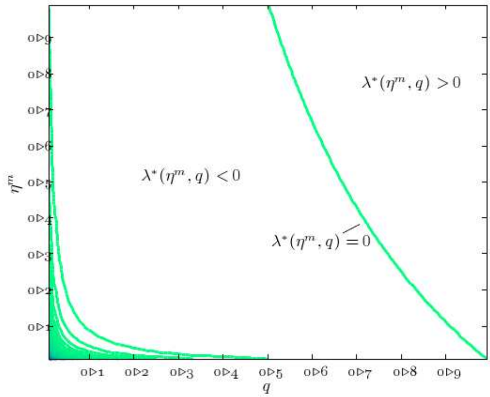


Figure 2.16: Set of admissible strategies for a politician who cares in reelection as a function of η^m and q. The function $\lambda^*(\eta^m, q)$ monotonically increases in η^m and q. On the right of the contour line where $\lambda^*(\eta^m, q) = 0$ the pure strategy of cheating is not admissible. On left side $0 \leq \lambda \leq 1$ is always greater than $\lambda^*(\eta^m, q)$ as it is now negative (see inequality (2.11)). Any combination of q and η^m in this area makes viable for the government to cheat citizens. Note that even though voters think that the state $\sigma = H$ occurs almost surely (i.e. for q very close to one), a distribution of education collapsed around zero allows the government to cheat them.

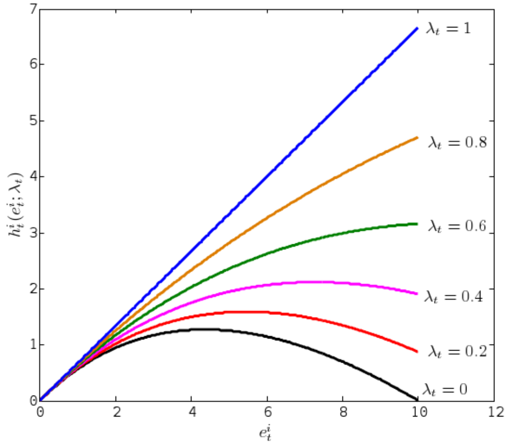
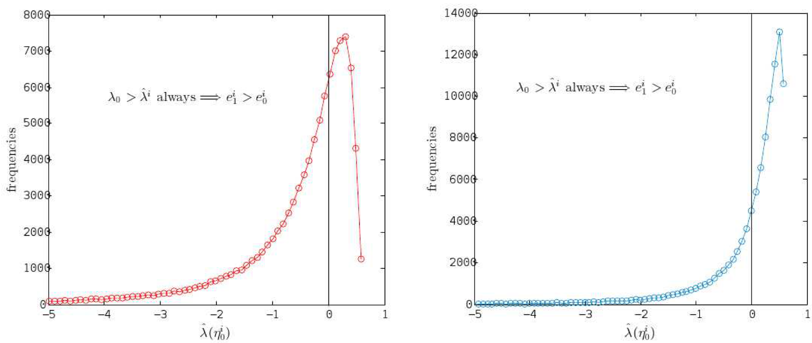


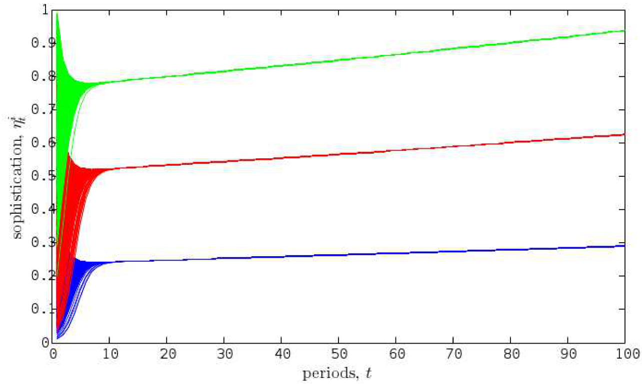
Figure 2.17: Private investments h_t^i as a function of human capital stock e_t^i for different values of ruler's congruence rate λ_t. Parameters are: $\delta = 0$, $\Omega = 0.4$, and $\bar{e} = 10$. The plots underlines the inelastic response of private investments with respect to politics (on the left side) for naive persons .



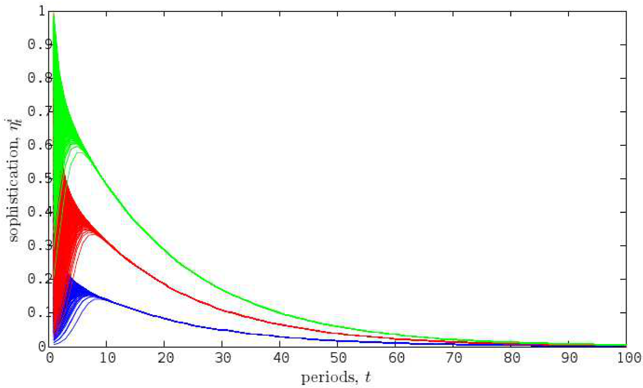
(a) *Skewed distribution,* $\eta_0 \sim Beta(2, 4)$.

(b) *Symmetric distribution,* $\eta_0 \sim Beta(2, 2)$.

Figure 2.18: $\tilde{\lambda}(\eta_0^i)$ distribution in a society of $n = 10^5$ citizens drawn with parameters $\delta = 0.2$, $\beta = 0.8$, and $\tau = 0.3$. Note that in a more skewed distribution more citizens are willing to invest in human capital.



(a) *Multiple steady-states* $(q \geq \bar{q}(\beta, \delta, \tau))$.



(b) *A unique steady-state* $(q < \bar{q}(\beta, \delta, \tau))$.

Figure 2.19: Different initial distributions of education follow different paths in panel (a) whereas converge to a unique steady-state equilibrium ($e_\infty = 0$) in panel (b). Each line describes the trajectory of an individual $i = 1, \ldots, 10^3$ for $t = 1, \ldots, 100$. Green lines draw a negative skewed distribution ($\eta_0 \sim Beta(6, 2)$), red lines a symmetric one ($\eta_0 \sim Beta(2, 2)$), and blue lines a positive skewed one ($\eta_0 \sim Beta(2, 6)$). Same colors describe same distributions at $t = 1$ among panel (a) and (b). Note that convergence is caught up firstly by the sophisticated agents (that for first know what is going on) whereas the last to catch up are the most naive citizens.

Bibliography

[1] Abramowitz A., Lanoue D.J. and Ramesh S. (1988), *Economic conditions, causal attributions, and political evaluations in the 1984 presidential election*, Journal of Politics, Vol. 50, No. 4: 848–63.

[2] Acemoglu D., Robinson J.A. (2000), *Why did the west extend the franchise? Democracy, inequality, and growth in historical perspective*, QJE, Vol. 115, No. 4: 1167-1199.

[3] Acemoglu D., Johnson S., Robinson J.A. and Yared P. (2005), *From education to democracy?*, AER, Vol. 95, No. 2: 44-49.

[4] Acemoglu D. (2005), *Politics and economics in weak and strong states*, JME, Vol. 52: 1199?1226.

[5] Acemoglu D., Egorov G. and Sonin K. (2010), *Political selection and persistence of bad governments*, QJE, Vol. 125, No. 4: 1511–1575.

[1] Acemoglu D., Egorov G., and Sonin K. (2013), *A Political Theory of Populism*, QJE, Vol. 128, No. 2: 771-805.

[7] Ashworth S. (2012), *Electoral accountability: recent theoretical and empirical work*, ARPS No. 15: 183-201

[8] Ashworth S., de Mesquita E.B. and Friedenberg A. (2013) *Accountability Traps*, Working Paper.

[9] Austen-Smith D., Banks J. (1989), *Elections, Coalitions, and Legislative Outcomes*, APSR, Vol. 82, No. 2: 405-422.

[4] Banks, J. (1990), *A model of electoral competition with incomplete information*, JET, Vol. 50, No. 2: 309-325.

[5] Banks, J., Sundaram R. (1993), *Adverse Selection and Moral Hazard in a Repeated Elections Model*, in Political Economy: Institutions, Information, Competition, and Representation, eds. W. Barnett, M.J. Hinich and N. Schofield. New York, Cambridge University Press.

[12] Barro R.J. (1973), *The Control of Politicians: An Economic Model*, PC, Vol. 14: 19-42

[13] Barro R. J. (1991), *Economic Growth in a Cross Section of Countries*, QJE, Vol. 106: 407-433.

[14] Barro R.J. (1999), *Determinants of democracy*, JPE, Vol. 107: 158-183.

[15] Barro R.J and Sala-I-Martin X. (1992), *Public Finance in Models of Economic Growth*, RES, Vol. 59, No. 4: 645-661.

[16] Benabou R. (1996), *Inequality and growth*, NBER Macroeconomics Annual 1996, Vol. 11.

[17] Benabou R. (2000), *Unequal societies: income distribution and the social contract*, AER, Vol. 90, No. 1: 96-129.

[18] Benabou R. (2002), *Tax and Education Policy in a Heterogeneous Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?*, Econometrica, Vol. 70, No. 2: 481-517

[6] Benabou R. (2013), *Groupthink: collective delusions in organizations and markets*, IZA Discussion Papers 7322, Institute for the Study of Labor.

[7] Benabou, R. and Tirole J. (1999), *Self-confidence: intrapersonal strategies*, IDEI mimeo.

[8] Benabou R., Tirole, J. (2002), *Self confidence and personal motivation*, QJE, Vol. 117, No. 3: 871-915.

[22] Benhabib J., Schmitt-Groh S. and Uribe M. (2001), *Monetary Policy and Multiple Equilibria*, AER, Vol. 91, No. 1: 167-186.

[23] Besley T., Case, A. (1995), *Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition*, AER Vol. 85, No. 1: 25-45.

[24] Besley T. (2006), *Principled Agents? The Political Economy of Good Government*, Oxford University Press.

[25] Bidner C. and Francois P. (2013), *The Emergence of Political Accountability*, QJE, Vol. 128, No. 3: 1397-1448.

[26] Bidner C., Francois P., and Trebbi F. (2014), *A Theory of Minimalist Democracy*, Working Paper.

[27] Bobba M., Coviello D. (2007), *Weak instruments and weak identification, in estimating the effects of education, on democracy*, EL, Vol. 96, No. 3: 301-306.

[28] Bourguignon F., Verdier T. (2000), *Oligarchy, democracy, inequality and growth*, JDE, Vol. 62: 285-313.

[29] Callender S. and Wilkie S. (2007), *Lies, damned lies, and political campaigns*, GEB, Vol. 60: 262-286.

[30] Castell-Climent A. (2008), *On the distribution of education and democracy*, JDE, Vol. 87: 179-190.

[31] Chappell H.W. and Keech W.R. (1985), *A new view of political accountability for economic performance*, APSR: 10–27.

[32] Coate, S., Morris, S. (1995), *On the Form of Transfers in Special Interests*, JPE, Vol. 103, No. 6: 1210-35.

[33] Durlauf S.N. (1996), *A Theory of Persistent Income Inequality*, JEG, Vol. 1: 75-93.

[15] Fearon, J. (1999), *Electoral Accountability and the Control of Politicians: Selecting Good Candidates versus Sanctioning Poor Performance*, in Democracy, Accountability, and Representation, eds. A. Przeworski, S.C. Stokes and B. Manin, New York, Cambridge University Press.

[35] Federico C.M. and Sidanius J. (2002), *Sophistication and the antecedents of Whites' Ratial-Policy Attitudes: Racism, ideology, and Affirmative Action in America*, Public Opinion Quaterly, Vol. 66: 145-176.

[16] Ferejohn J. (1986), *Incumbent Performance and Electoral Control*, PC, Vol. 50, No. 1/3, Carnegie Papers on Political Economy, Volume 6: 5-25.

[37] Galor O. and Zeira J. (1993), *Income Distribution and Macroeconomics*, RES, Vol. 60, No. 1: 35-52.

[38] Galor O. and Moav O. (2006), *From physical to human capital accumulation: Inequality and the process of development*, RES Vol. 71, No. 4: 1001–1026.

[39] Galor O. (2011), *Inequality, human capital formation and the process of development*, NBER Working Paper.

[40] Glaeser E.L, La Porta R., Lopez-de-Silanes F., Shleifer A. (2004), *Do institutions cause growth?*, JEG, Vol. 9, 271-303.

[41] Glaeser E.L. and Shleifer A. (2005), *The Curley effect: The economics of shaping the electorate*, JLEO, Vol. 21, No. 1: 1–19

[18] Glomm, G. and Ravikumar, B. (1992), *Public versus private investment in human capital endogenous growth and income inequality*, JPE, Vol. 100(4) :818-34.

[19] Gomez B.T. and Wilson J.M. (2001), *Political sophistication and economic voting in the American electorate: A theory of heterogeneous attribution*, AJPS: 899–914.

[20] Gradstein M. and Justman M. (1997), *Democratic choice of an education system: implications for growth and income distribution*, JEG, Vol. 2, No. 2: 169–183.

[45] Grossman G.M. and Helpman E. (2001), *Special interest politics*, The MIT press, Cambridge, Massachusetts.

[46] Hall R.E. and Jones C.I. (1999), *Why Do Some Countries Produce So Much More Output Per Worker Than Others?*, QJE, Vol. 114, No. 1: 83-116.

[21] Harrington, J.E. (1993), *Economic Policy, Economic Performance, and Elections*, AER, Vol. 83, No. 1: 27-42.

[48] Herbst J.I. (2000), *States and power in Africa: comparative lessons in authority and control*, Princeton University Press, Princeton.

[49] Kartik N. and McAfee R.P. (2007), *Signaling Character in Electoral Competition*, AER, Vol. 97, No. 3: 852-870.

[50] Kaufmann D., Kraay A. and Mastruzzi M. (2010), *The Worldwide Governance Indicators: Methodology and Analytical Issues*, World Bank Policy Research Working Paper No. 5430.

[51] Knack S. and Keefer P. (1995), *Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Indicators*, Economics and Politics, Vol. 7, No. 3: 207-228.

[52] Lipset S.M. (1959), *Some social requisites of democracy: economic development and political legitimacy*, APSR, Vol. 53, 69-105.

[53] Lucas, R J. (1988), *On the mechanics of economic development*, JME, Vol. 22(1): 3-42.

[54] McGraw K.M. (2000), *Contributions of the cognitive approach to political psychology*, Political Psychology, Vol. 21, No. 4: 805–832.

[55] Migdal J. (1988), *Strong societies and weak States: State-society relations and State capabilities in third world*, Princeton University Press, Princeton.

[56] Przeworski A., Alvarez M., Cheibub J.A., Limongi F. (2000), *Democracy and development: political institutions and material well-being in the world, 1950-1990*, Cambridge University Press, NY.

[57] Persson T. and Tabellini G.E. (2000), *Political economics: explaining economic policy*, MIT Press, Cambridge, Massachusetts.

[58] Rogoff, K. (1990), *Equilibrium Political Budget Cycles*, AER, Vol. 80, No. 1: 21-36.

[59] Rogoff K. and Sibert A. (1988), *Elections and macroeconomic policy cycles*, RES, Vol. 55, No. 1: 1–16.

[60] Saint-Paul G. (1994), *Unemployment, wage rigidity, and the returns to education*, EER, Vol. 38, No. 3: 535–543.

[61] Sniderman P., Brody R.A. and Tetlock P.E. (1991), *Reasoning and Choice: Explorations in Political Psychology*, Cambridge: Cambridge University Press.

[62] Zallen J. (1992), *The nature and origins of mass opinion*, Cambridge university press.

Chapter 3

A Political Psychology Theory Of Populism

Why is it that certain countries keep repeating the same 'mistakes' and never learn? In fact, once the political and institutional incentives and constraints are taken into account, policies that appear to be mistakes are perfectly rational responses to distorted or imperfect political incentives.

Alesina A. (1988)

There has been an increasing attention to populism in Political Economy in the last years. Economic theories attempting to explain populism have mostly focused on politician's incentive to conform to popular opinion. Starting from this literature (Harrington, 1993; Canes-Wrone et al., 2001; Chiu, 2002; Heidhues and Lagerlöf, 2003; Maskin and Tirole, 2004), Frisel (2009) has argued that populist politicians are the ones that always pander to the median voter's preferences, regardless of the implications for the social welfare. Hodler, Loertscher, and Rohner (2010) went further by arguing the incumbent uses inefficient policies to increase the information asymmetry and improve his chances of reelection. More recently Acemoglu, Egorov, and Sonin (2013) develop the idea, originally proposed by Dornbush and Edwards (1991) in the field of Latin America, that an honest politician chooses populist leftist policies as a way of signaling that he is not beholden to the interests of the right.

The Latin American countries experience has been broadly studied in Political Science, since it rose an unresolved political puzzle according to which populist politicians has been widely supported by the electorate while ultimately hurt the economic interests of the majorities[1]. However, in the last decades new forms of populism have appeared, especially in Europe, and the increasing popularity of these parties are now presented as the main threat to our democracy (Panizza, 2005), behaving in a demagogic fashion and recklessly claiming big change for the better are possible and that they can make them happen (Albertazzi and McDonnell, 2008).

The rise of the new populism in Europe has been favored by what Dahrendorf has arguably called the erosion of the 'class'of politic (Dahrendorf, 1996). Along with significant changes in the underlying social structure, firstly and most importantly decades of political and economic stability, advanced countries have been experienced the emergence of a new political class, to who Dahrendorf refers as 'mediocre', elected by 'mediocre people'as a result of a changing in the relationship between civil society and political institutions. This work attempts to model such relationship and to explain the emergence and persistence over time of the populism phenomenon in particular countries (and not in others) by originally looking at the electorate side that, in different societies, may be more or less sophisticated. The distribution of political sophistication within a country generates different incentive structure for the incumbent that accordingly optimally decide whether to be a populist or a responsible type whereas between countries might determine completely different equilibria in the long run, one with populist politicians and one ruled by responsible ones. I argue that selfish politicians have the chance to behave in a populist fashion when a naive electorate fail in keeping rulers politically accountable. Naive voters are basically unaware of the politicians intentions providing to the latter opportunities for the manip-

[1]The political science literature on populism is vast. Notable works include Dornbush and Edwards (1991), Knight (1998), Edwards (2010), Levitsky and Roberts (2011), Connif et al. (2012).

ulation of the economy and the electoral outcome, claiming that investment are viable in order to increase his expected future rents by increasing the mean of the human capital distribution (which determines the extent of the tax revenues) and, at the same time, by introducing more spread (inequality) in the distribution. Such a strategy, carried out by populist rulers, aimed to maximize electoral consensus, by enlarging the naive segment of the society, and expected future rents by making the rich segment of the society even richer.

To address the importance of information and communication in politics, we propose a dynamic signaling political model where citizens/voters are endowed by different level of education and, according to his own level of education, each of them codifies the signal (the announcement on public investment) sent by the ruler *differently* (see Bénabou and Tirole, 2002). Despite the signal is costless, it still conveys information to the electorate who may guess investment are viable when actually they are utterly inefficient. Despite citizens do prefer more public investments that increase individual productivity, they are politically committed to responsible economic policy never allowing politicians in office to run investments whose costs exceed the individual benefit. Once it is assessed that the ruler is indeed a populist type, citizens react by voting him out and replacing him with a challenger. Such level of awareness is never achieved by naive citizens who accordingly never oppose to a reelection of a populist ruler. Therefore, populist politicians build their consensus on the naive segment of the society that, being unaware, results to be easily manipulable by the former ones. Once elected, their actions ultimately advantage the rich who anticipate he is reckless and, despite the political commitment, responds by carrying out more investment in human capital. At the end of day, any *political* commitment to keep economic policy responsible turns to be *economically* non credible.

This mechanism outlines non negligible differences in the long run among different society endowed by different initial education distribution. Politicians increase their rents by ruling over naive and unequal societies, that is one in which the median is naive enough to allow political manipulation and the mean is high enough to increase expected future tax revenues. The initial distribution of education is crucial to determine the mechanisms of such a process, shaping the economic/political development path of a country. If the bad state of the world is persistent enough, this process may generate long run equilibria characterized by poverty traps, that involve the naive segment of the society, and high inequality. Fortunately, this is not expected to happen in sophisticated societies. We show how more sophisticated society are more able to account for populist politicians that in turn are less likely to lie. As a result sophisticated society converge to an equilibrium with less inequality and higher incomes.

The fact that a minoritarian part of the electorate, the élite, gets better with populist policies is a milestone of the literature on populism. Arguably, this suddenly arises a puzzle on the political sustainability over time of such a policy, that has not very clearly discussed in the aforementioned literature, and still presents in all the countries involved, starting from the America Latina countries. Why political support for keeping populist rulers accountable does not arise? Why should the majority bring support to a populist ruler if, in the long run, he leads them to poverty? The answer here proposed, using tools of political psychology, is that the naive majority is not very well aware of politicians intentions and accordingly it is not prepared to fight for. As long as the government budget constraint binds, their naiveté persists and broadens over time getting room for rulers to behave (increasingly) in a populist fashion.

Populism is therefore of primary urgency in not well functioning democracies where the electorate is, to a great extent, naive that are more prone to believe to populist announcements and to be cheated by selfish politicians. We propose a novel explanation according to which more educated societies are more able to punish politicians that, in turn, only invest in public goods when it is responsibly possible. The combination of the accumulation and political mechanism creates the potential for multiple steady states, one for societies with low education and populist rulers and one for societies with high education and responsible rulers.

This chapter is organized as follows. Despite the main features of the model are basically invariant with respect to the model introduced in Chapter 2, in Section 3.1 we shall discuss the way populism affects voters beliefs, and we shall introduce the voting rules. The Markov Perfect Bayesian Equilibrium will be characterized in Section 3.2, firstly discussing how the accumulation process is affected by the institutional setting and then endogeneizing political choices as best responses of the sophistication rate of the electorate (via accountability). Section 3.3 discusses the dynamic implication of the political economic interaction and multiple equilibria are characterized. Section 3.4 concludes.

3.1 The model

3.1.1 Information and beliefs

In Chapter 2 we have considered the rulers incentives of lying the electorate when things go well, i.e. in state H, convincing them that no public investment are viable and pocketing what is left of the tax revenues by himself. In state L, nothing happened and citizens know that there is no room for the ruler to behave strategically, regardless of the level of sophistication. In reality, the incumbent may have the incentive to claim that investment are rather efficient (i.e. H-type investment) in order to expected future rents. This symmetric scenario allow us to study the consequences of a reckless incumbent politician that makes inefficient public investment even in state L.

We define this attitude of politicians as populism. Politicians are reckless with rate $1-\gamma_t$ and, according to his own level of sophistication, each citizen i guesses that the incumbent ruler is following populist policies with probability (or reckless rate)

$$P(\hat{\sigma}_t^i = H | \sigma_t = L) = (1 - \gamma_t)\eta_t^i \quad (3.1)$$

where $\gamma_t = 0$ stands for a fully populist ruler and $\gamma_t = 1$ for responsible type. However we allow γ_t to range as a continuum from 0 to 1. With

$$P(\hat{\sigma}_t^i = L | \sigma_t = L) = 1 - (1 - \gamma_t)\eta_t^i \quad (3.2)$$

the ruler is rather thought to be responsible, always signaling the right state. As one may see, fully naive citizens, with $\eta_t^i = 0$, always believe that the ruler is a congruent type always signaling the right state of the world independently from his true type, i.e. an admissible value of γ_t. To keep things easy we impose throughout $\lambda_t = 1$ such that in state H public investments are always provided. As in Chapter 2 we allow each citizen to infer the probability that the ruler is cheating, signaling H when actually the true state is L by using the following awareness-management model à la Bénabou-Tirole:

$$\begin{aligned} p_t^i(\gamma_t, \eta_t^i, q) &\equiv P(\sigma_t = L | \hat{\sigma}_t^i = H; \gamma_t, \eta_t^i) \\ &= \frac{(1-q)(1-\gamma_t)\eta_t^i}{q + (1-q)(1-\gamma_t)\eta_t^i} \end{aligned} \tag{3.3}$$

3.1.2 Timing of events

The timing of events within every period is as follows:

T1 Nature draws $\sigma_t = \{H, L\}$, that is private information of the ruler. Each citizen inherits e_t^i from the private investment made at time $t-1$ and benefits from the public investment made by the former government, G_{t-1}.

T2 Politician in office chooses the action γ_t, the rate of economic policy responsibility, and, accordingly, invest in a public good (that will be productive in $t+1$), i.e. chooses $G_t = \{0, 1\}$.

T3 Citizens plan to invest h_t^i in human capital based on their beliefs on the ruler's type and, simultaneously, elections are held (the median citizen chooses $\varphi_t = \{0, 1\}$ based on posterior beliefs).

T4 Payoffs are given by rents and consumptions to politician and citizens respectively.

3.2 The Markov Perfect Bayesian Equilibrium

3.2.1 Elections

Citizens vote retrospectively according to the evidence they have collected on political announces. Since the government's strategy is realized only after the elections, only individual beliefs are involved in the inferential process. Every citizen i processes all the informations collected and votes again for the incumbent if she has no evidence of the fraud, i.e. iff the evidence E_t in favor of the hypothesis p_t^i is not positive:

$$E(p_t^i) \equiv \log\left(\frac{p_t^i}{1-p_t^i}\right) \leq 0 \tag{3.4}$$

which occurs where $p_t^i \leq 1/2$. It in turn means that if the majority of them has no evidence about the cheating move of the policymaker he will be reelected, contingency that occurs when $P(p_t^i \leq 1/2) = F(1/2) \geq 1/2$. Now, given $0 \leq p_t^m \leq q$ with $F(p_t^m) = 1/2$, we require that

$$F\left(\frac{1}{2}\right) \geq \frac{1}{2} = F(p_t^m) \Longleftrightarrow p_t^m \leq \frac{1}{2} \tag{3.5}$$

by monotonicity of $F(\cdot)$. In other words, it turns out that the policymaker won't be reappointed if the median citizen thinks that he is plausibly cheating them. Therefore, if politician cares about reelection he would be willing to push down p_t^m at least to $1/2$. This is of course easier in a society where people can easily be made fools, i.e. in one with a skewed distribution of η_t^i. We summarize this result in Proposition 1:

PROPOSITION 1: *Let* $p_t^m(\gamma_t, \eta_t^m, q) \equiv P(\sigma_t = L|\hat{\sigma}_t^m = H; \gamma_t, \eta_t^m)$.

(i) *If* $p_t^m(\gamma_t, \eta_t^m, q) \leq \frac{1}{2}$ *the optimal strategy in the stage game is to play* $\varphi_t = 1$.

(ii) *If* $p_t^m(\gamma_t, \eta_t^m, q) > \frac{1}{2}$ *the incumbent will not be retained (i.e.* $\varphi_t = 0$*).*

Substituting (3.3) into (3.5), we get the set of admissible strategies for a policymaker who cares about reelection:

$$\gamma_t \geq 1 - \frac{1}{\eta_t^m}\frac{q}{1-q} \equiv \gamma_t^*(\eta_t^m, q) \tag{3.6}$$

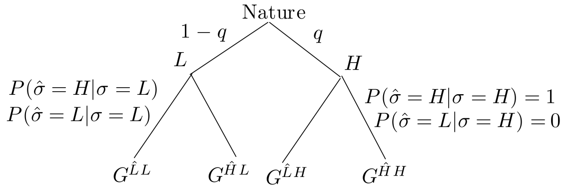


Figure 3.1: Public investments subgame. Note that when the state is thought to be strong every citizens know that the ruler will play $G_t = G^{H\hat{H}} = 1$, no matter the educational level of him. On the contrary, when the state is thought to be weak public investments can be positive ($G^{L\hat{H}} = 1$) or null ($G^{L\hat{L}} = 0$) with probabilities given respectively by equations (3.1) and (3.2).

such that the optimal strategy of the ruler is to play

$$\max\Big[0, \gamma_t^*(\eta_t^m, q)\Big] \leq \gamma_t \leq 1. \tag{3.7}$$

It is worth noting that $\partial\gamma_t^*/\partial\eta_t^m \geq 0$, i.e. that a more sophisticated society force the incumbent to be more responsible. This is what I mean by sophisticated accountability.

3.2.2 The accumulation process under inefficient projects

In period T3, citizens, along with voting, invest in human capital based on their beliefs on the ruler's type. According to Figure 3.1, politicians are thought of to invest in state L using the following mixed strategy:

$$G_t^L = \begin{cases} 1 & \text{with} \quad (1-\gamma_t)\eta_t^i \\ 0 & \text{with} \quad 1-(1-\gamma_t)\eta_t^i \end{cases} \tag{3.8}$$

whereas $G_t^H = 1$ with certainty.

Investment in state L involve only inefficient projects characterized by a fraction κ^L of unproductive returns. The rate of inefficiency of a public project is therefore a contingent state parameter that takes values $\kappa^H = 0$ and $\kappa^L > 0$.

ASSUMPTION 1: *Let* $\bar{e}_{t+1} = \max(e_{t+1}^i)$ *and* B *the cost of the project.*

$$\kappa^L \geq \frac{\mathbb{E}[\bar{e}_{t+1}]^2 - B}{\mathbb{E}[\bar{e}_{t+1}]^2}. \tag{3.9}$$

Assumption 1 says that social costs coming from inefficient projects exceed individual benefits[2]. When Assumption 1 holds even the richest citizen individually gets less than what the government spends for the inefficient project. We now introduce a further assumption aimed at allowing citizens to internalize this social cost.

ASSUMPTION 2: *Citizens are politically committed to responsible economic policy.*

Assumption 2 is justified from the fact that rational citizens look forward to the catastrophic consequences to follow populist policies over time anticipating that in the future there might be greater losses[3]. Under Assumptions 1 and 2 any politician that is believed to be reckless will not be retained. This is because citizens internalize the social cost that comes from investing in an inefficient project.

Note that Assumption 2 introduces a *political* commitment but nothing says on the rational economic behavior that each citizens will take in time T3. Note further, from equation (3.4), that naive citizens fail in collecting all the information to be aware of the politician intentions so as, despite all the citizens potentially are politically committed to economic responsibility, they might, in absence of any evidence, reelect the populist type.

It follows that the expected future individual output will be

$$\begin{aligned}\mathbb{E}_t(y^i_{t+1}) &= \Big[P(\hat{\sigma}^i_t = H|\sigma_t = H)P(\sigma_t = H) \\ &\quad + P(\hat{\sigma}^i_t = H|\sigma_t = L)P(\sigma_t = L)(1-\kappa)\Big](e^i_{t+1})^\alpha \\ &= \Big[q + (1-q)(1-\kappa)(1-\gamma_t)\eta^i_t\Big](e^i_{t+1})^\alpha\end{aligned} \tag{3.10}$$

[2]One may see that by comparing the expected utility that each citizen i will take in state L under inefficient investment, $(1-q)(1-\kappa)(1-\gamma_t)\eta^i_t\mathbb{E}[e^i_{t+1}]^2$, with the correspondent expected social costs, $(1-q)(1-\gamma_t)B$. Assumption 1 amounts to state that $\eta^i_t\mathbb{E}[e^i_{t+1}]^2 \leq B + \kappa\eta^i_t\mathbb{E}[e^i_{t+1}]^2$ for each i.

[3]One way to do so is to introduce a negative drift, $\nu(\gamma_t)$, in the law of motion of $q_{t+1}(q_t)$, characterized in Section 2.7.1, increasing in the populist attitude of the ruler. In this fashion,

where $(1-\kappa)(1-\gamma_t)\eta_t^i(e_{t+1}^i)^\alpha$ are informational rents of sophisticated citizens that occur when the state of the world is L, the politician in office is not fully responsible, and $\kappa < 1$.

Recursively, each citizen maximizes the expected current period return that will be consumed according to (2.4) and the agent i's intertemporal utility at time t is

$$V(e_t^i) = \max_{h_t^i} \Big\{ (1-\tau)y_t^i - \frac{1}{\phi}(h_t^i)^\phi + \beta q \Big[\mathbb{E}_t\big[V^H(e_{t+1}^i)|\hat{\sigma}^i = H\big] + \\ + \mathbb{E}_t\big[V^H(e_{t+1}^i)|\hat{\sigma}^i = L\big]\Big] + \beta(1-q)\Big[\mathbb{E}_t\big[V^L(e_{t+1}^i)|\hat{\sigma}^i = H\big] + \\ + \mathbb{E}_t\big[V^L(e_{t+1}^i)|\hat{\sigma}^i = L\big]\Big]\Big\} \tag{3.11}$$

where current output are given by (2.14). Maximization yields the optimal effort taking in by each citizen i as a function of the state variable of the economy, e_t^i, and political institution, γ_t.

PROPOSITION 2: *Let $\Gamma \equiv 2\beta(1-\tau) \leq 1$. In every period t, each citizen i optimally chooses a level of investment increasing with the populist attitude of the ruler. The reaction to a populist announcement is as strong as larger is the level of sophistication η_t^i.*

$$h_t^i(e_t^i;\gamma_t) = (1-\delta)\left[\frac{\Gamma(q+(1-q)(1-\kappa)(1-\gamma_t)\eta_t^i)}{1-\Gamma(q+(1-q)(1-\kappa)(1-\gamma_t)\eta_t^i)}\right]e_t^i \tag{3.12}$$

When the state is $\sigma_t = L$ the need for responsible politicians ruling public business is more subtle. A responsible ruler is indeed a brake for the accumulation process of human capital, at least in the short term, and producers react to higher values of γ_t decreasing private investment, i.e. $\partial h_t^i/\partial\gamma_t \leq 0$. The question of who is the perfect agent suddenly arises but such a dilemma only involves sophisticated citizens who, while politically committed to responsible economic policy, take advantage from this

the law of motion of q_t becomes $q_{t+1} = (1-\xi) - \nu(\gamma_t) + (2\xi - 1)q_t$ and higher values

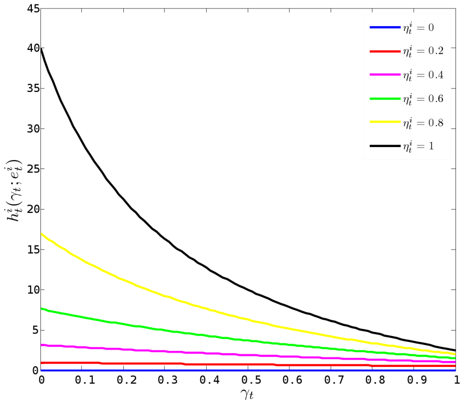


Figure 3.2: Private investments h_t^i as a function of human capital stock e_t^i for different values of ruler's rate of responsability γ_t. Parameters are: $\delta = 0$, $\Omega = 0.4$, $q = 0.5$, $\kappa = 0$, and $\bar{e} = 10$. The plots underlines the inelastic response of private investments with respect to politics (on the left side) for naive persons.

policy in an increasing fashion to his level of awareness ($\partial^2 h_t^i / \partial\gamma_t \partial e_t^i \leq 0$), as Figure 3.2 shows.

Exogenous parameters also shape the optimal investment effort. For example, taxation (τ) discourage private investments as well as a greater depreciation (δ) of human capital and a greater extent of unproductive returns of the investment (κ). Conversely, regardless the state σ_t, more patient agents (β) are more willing to invest in order to consume more in the future. Finally, an higher belief on the occurrence of the state H makes investments going up (i.e. $\partial h_t^i / \partial q \geq 0$). However, thanks to populist policies, even when $q = 0$ citizens are still encouraged to invest in human capital and the investment is found to be monotonically increasing with e_t^i.

Looking at the motion of the human capital

$$e_{t+1}^i(e_t^i; \gamma_t) = \frac{1 - \delta}{1 - \Gamma(q + (1 - q)(1 - \kappa)(1 - \gamma_t)\eta_t^i)} e_t^i \tag{3.13}$$

---

of $\nu(\cdot)$ make the state L more likely to occur in the future. To keep things easy, we assume that $\xi = 1$ and $\nu(\gamma_t) = 0$.

that differently from the rent seeking case (see Chapter 2)[4], is found to be convex with e^i_t. The consequences of the convexity of the state equation are striking and they are summarized in Proposition 3.

PROPOSITION 3: *Let* $\Omega \leq 1$.

(i) If q *is sufficiently small* ($\delta \geq \Omega$)*, there is only one stable steady state equilibrium represented by* $e_\infty = 0$.

(a) Naive citizens with a human capital level below

$$e_\infty(\gamma_\infty) = \frac{\delta - \Omega}{\Gamma(1-q)(1-\kappa)(1-\gamma_\infty)}\bar{e} \tag{3.14}$$

converge to a zero level of education ($e_\infty = 0$)*.*

(b) Richest citizens with $e^i_0 > e_\infty(\gamma_\infty)$ *accumulate over time taking advantage of populist announcements.*

(ii) If q *is sufficiently high* ($\Omega \geq \delta$)*, productive public investment are more likely and society (but the unskilled) accumulate over time getting more educated.*

Conversely with the rent-seeking case, here all the insights of model are brought up from the state $\sigma_t = L$. First of all, the model predicts that richest citizens (and only them), i.e. those with $e^i_0 > e_\infty(\gamma_\infty)$, take advantage of populist politicians, getting reacher and reacher over time. Although, the fact that a minoritarian part of the electorate gets better with populist policies is a milestone of the literature on populism, this suddenly arises a puzzle on the political sustainability over time of such a policy, not very clearly discussed in the aforementioned literature, and still present in all the countries involved, starting from the America Latina countries[5]. As the model predicts, if $e^m_0 < e_\infty(\gamma_\infty)$, the majority

[4] In Chapter 2 κ is set to be 0.

[5] While initially the strongest experiments of populism in Latin America obtained remarkable economic goals on the side of the budget deficit, nominal wages, growth, and in fostering industrialization (Kauffman and Stalling, 1991), they sooner or later run into serious external bottleneck that led these countries to economic and political collapse. This

of the society experiments, in the long run equilibrium, an indirect loss from such a policy. So why political support for keeping populist rulers accountable does not arise? Why should the majority bring support to a populist ruler if, in the long run, he leads them to poverty? The answer here proposed is that the naive majority is not very well aware of politicians intentions and accordingly it is not prepared to fight for and, as long as $\delta \leq \Omega$, their naivete persists and broadens over time getting room for rulers to behave (increasingly) in a populist fashion.

Turning back to the role of political institutions, it is worth noting that $\gamma_\infty \to 1$ pushes the unstable equilibrium $e_\infty(\gamma_\infty \to 1) \to \infty$ slowing the speed of convergence to $e_\infty = 0$, waiting for better moments. Things are persistently bad, and in the long run a high probability of having $\sigma_t = L$ leads, despite the responsibility of the polity, the society (and all the citizens there in) to poverty. However, the slowness of the convergence may increase the likelihood that some external shocks move the society toward another development path (turning σ_t from L to H).

3.2.3 The political process

We now turn on the political process through which sophisticated citizens bind reckless politicians. This allows us to explain away why only some country deals with populism whereas others do not.

In T2 the ruler anticipates what is the level of private investments made by each citizen i and, accordingly, chooses the optimal rate of political responsibility $\gamma_t = P(\hat{\sigma}_t = L|\sigma_t = L)$, ranged according to (3.7). After that, the politician decides whether or not to invest in the public good with probability γ_t, that is:

$$G_t^L = \begin{cases} 1 & \text{with} \quad 1 - \gamma_t \\ 0 & \text{with} \quad \gamma_t \end{cases} \tag{3.15}$$

The problem of the ruler turns to be a constrained maximization prob-

---

has been the case for example of Salvador Allende in Chile (1970-73), of Juan Peron in Argentina (1973-76), and Alan Garcia in Peru (1985-90), among others.

lem of the following kind:

$$\max_{\gamma_t} \quad V_t^r(\gamma_t) = T_t - [q + (1-q)(1-\gamma_t)]B + \beta \mathbb{E}_t\big[V_{t+1}^r(\gamma_t)\big] \quad \text{s.t. (3.7)} \tag{3.16}$$

The public good will be implemented either if the state is H or if the state is L and the ruler is not completely responsible. Overall, the expected cost is $[q + (1-q)(1-\gamma_t)]B$ where $B = bT_t$, with $b \in [0,1]$ by imperfect credit markets. Political decisions are also expected to influence future rents, in case of reelection, by determining the future (average) income of the economy, m_{t+1}, via the accumulation process. It follows that m_t is exactly identical to the rent-seeking case, but not the expected future human capital, μ_{t+1}, and wealth that are obtaining, respectively, from equation (3.13) and (3.10)[6]:

$$\mu_{t+1} = \mu_t + \frac{\Delta_t^2}{2} + (\Omega - \delta) - \Gamma(1-q)(1-\kappa)(1-\gamma_t)\eta_t \tag{3.17}$$

$$m_{t+1} = 2\mu_t + \Delta_t^2 + 2(\Omega-\delta) - 2\Gamma(1-q)(1-\kappa)(1-\gamma_t)\eta_t + \frac{1-q}{q}(1-\kappa)(1-\gamma_t)\eta_t \tag{3.18}$$

with $\eta_t = \exp(\mu_t + \Delta^2/2)/\bar{e}$. The problem (3.16) therefore can be reexpressed as

$$\begin{aligned} \ln V_t^r(\gamma_t) = \max_{\lambda_t} \Big\{ & \ln\tau + 2\mu_t + \Delta_t^2 - [q + (1-q)(1-\gamma_t)]b \\ & + \beta \mathbb{E}_t\big[\ln V_{t+1}^r(\gamma_t)\big]\Big\} \quad \text{s.t. (3.7)} \end{aligned} \tag{3.19}$$

Maximizing rents amounts to choice an optimal populism rate, ranged according to (3.7). By doing that the incumbent trades off future tax revenues with current rents coming from greater (unsustainable) public investment. Due to the linear functional form of rents, it is easy to note that $\partial \ln V_t^r/\partial\gamma_t \leq 0$ iff

$$b \leq \beta\left(\frac{1-2\Omega}{q}\right)(1-\kappa)\eta_t \equiv \underline{b}(\beta, q, \tau, \kappa, \eta_t) \tag{3.20}$$

with $\eta_t \equiv \exp(\mu_t + \Delta_t^2/2)/\bar{e}$. In other words, rents are found to be de-

[6]Proofs are gathered in the appendix.

creasing with his rate of being responsible provided that the cost of the public investment, relative to tax revenues, is small enough. Interestingly, the threshold $\underline{b}$ is found to be increasing with β, q, $-\tau$, and $-\kappa$, meaning that better economic conditions rise the incumbent's incentives to behave in a populist fashion. When condition (3.20) holds, incumbent rulers maximize office rents by pushing down γ_t as much as they can. However, given the accountability effort exerted by voters, the lowest still optimal value is the maximum between $\gamma_t^* = 1 - (1/\eta_t^m)(q/(1-q))$ and zero, according to (3.7). We summarize this result in Proposition 4:

PROPOSITION 4: *Political equilibrium. Assume* $\ln e_0^i \sim \mathcal{N}(\mu_0, \Delta_0^2)$. *There exists* $\underline{b}$ *and* $\bar{b}$, *such that:*

(i) If $b \leq \underline{b}$ *the incumbent plays* $\gamma_t = \gamma_t^*(\eta_t^m)$ *and the median voter plays* $\varphi_t = 1$.

(ii) If $b \in (\underline{b}, \bar{b}(\delta))$ *the incumbent plays* $\gamma_t = 1$ *and the median voter plays* $\varphi_t = 1$.

(iii) If $b \geq \bar{b}(\delta)$ *the incumbent plays* $\gamma_t = 0$ *and the median voter plays* $\varphi_t = 0$ *(Go-for-Broke).*

When public investments are not so costly the incumbent takes the electoral advantage against the opponent behaving as a populist politician by pushing the rate γ_t to be the lowest possible value. However, in a sophisticated society he must be responsible enough and, as a result, the optimal rate of responsibility is:

$$\gamma_t^* = 1 - \frac{1}{\eta_t^m}\left(\frac{q}{1-q}\right) \tag{3.21}$$

that is decreasing in q but increasing in the overall society's level of sophistication.

The cost of the public good drives political decisions for a given distribution of education in the society. When b is higher than $\underline{b}$ (i.e. $b > \underline{b}$) there is no incentive to cheat the electorate because investing recklessly is no longer sustainable as greater as the future discount factor increases ($\partial \underline{b}/\partial \beta \leq 0$). As Figure 3.3 points out there is a cut-point $\bar{b} \geq \underline{b}$ such

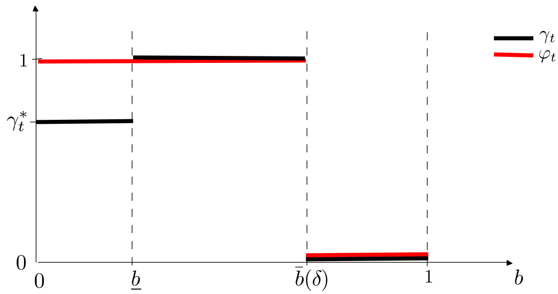


Figure 3.3: Political Equilibrium as a function of b.

that the cost of the project, while feasible (i.e. $b \leq 1$), could be so high to incentive the ruler to go-for-broke. If $b \geq \bar{b}$ playing go-for-broke by extracting all the tax revenues strictly dominates $\gamma_t = 1$. The median voter anticipates that for every $b > \bar{b}$ the incumbent would Go-for-Broke optimally responding by assigning probability zero on the event that the ruler plays $\gamma_t = 1$. Consequently, the median voter plays $\varphi_t = 0$ and the ruler, that anticipates this move, goes-for-broke.

In the Appendix we also show that $\bar{b}(\delta)$ is a decreasing function of the depreciation rate of the human capital of the producers so as for high level of δ Go-for-Broke is more likely to be the optimal strategy in the stage game. In particular, there exists a threshold δ^* such that for any $\delta < \delta^*$ Going-for-Broke is not an admissible strategy for the ruler, i.e. $\bar{b}(\delta) > 1$.

From now on we shall entirely focus on the more interesting case where $b \leq \underline{b}$.

3.2.4 The characterization of the MPBE

Once γ_t, and then G_t, are realized citizens come to know what is their pay-off, that is how much they have really invested and what is their production and what will be their human capital level in next periods:

$$h_t^i(e_t^i; q, e_t^m) = (1-\delta)\left[\frac{\Omega(1+(1-\kappa)\frac{e_t^i}{e_t^m})}{1-\Omega(1+(1-\kappa)\frac{e_t^i}{e_t^m})}\right]e_t^i \tag{3.22}$$

$$e_{t+1}^i(e_t^i; q, e_t^m) = \frac{1-\delta}{1-\Omega(1+(1-\kappa)\frac{e_t^i}{e_t^m})}e_t^i \tag{3.23}$$

Conversely with the rent-seeking case, all the MPBE values are now decreasing with the general level of sophistication of the society, e_t^m. A more sophisticated pivotal voter is indeed now more aware of the long-term drawbacks of public investment therefore acting as a brake for the short-term accumulation process. This positive external (long-term) effect, discussed in the subsequent subsection, works through the accountability effort pushed by the median citizen, that makes more likely a fair political environment.

Overall, the MBP equilibrium can be summarized in following Proposition:

PROPOSITION 5: *Assume* $\ln e_0^i \sim \mathcal{N}(\mu_0, \Delta_0^2)$*, there exists a unique Markov Bayesian Perfect Equilibrium such that*

(i) If the state is strong ($\sigma_t = H$) the incumbent ruler does invest ($G_t = 1$). Citizens know that and respond by investing $h_t^i = (1-\delta)\Omega/(1-\Omega)$, according to the parameters that govern the economy and regardless of the education they have; future stocks of human capital are then in equilibrium equal to $e_{t+1}^i = (1-\delta)e_t^i/(1-\Omega)$.

(ii) If the state is weak ($\sigma_t = L$) the incumbent ruler has the incentive to cheat citizens that in turn bind politicians by accountability. The ruler cares to be reappointed and the optimal strategy is given by (3.21) when condition (3.20) holds. Citizens, on the other side, guess the ruler does invest ($G_t = 1$) with probability $(1-\gamma_t)\eta_t^i$ and respond by investing and accumulating according to (3.22) and (3.23), respectively.

3.3 Dynamics and multiple steady states

The initial distribution of human capital strongly shapes the dynamics of the economy, political choices, and agents incentives. Right tailed distributions are mostly composed by naive agents who barely account for ruler's duties that, in turn, are allowed to implement populist polity (by playing small values of γ_t). However, these polities ultimately turn to have, in the long-run, atrocious effects for the former. Sophisticated agents (or the elite) instead take advantage from populist claims getting reacher and reacher. The result of a long-lasting period characterized by the impossibility of making productive public investment is therefore the rising of the inequality when populist politicians have the chance to govern the country. This is very implausible in highly sophisticated countries where sophisticated electoral accountability imposes to the ruler high values of γ_t as the price for reappointment. However, persistent crises may reduce the general level of sophistication of the society and encourage rulers to behave recklessly for electoral purposes.

The long-run equilibrium of the economy is then strictly depending on the initial distribution of the education and it is found as the solution of the following system which describes the joint evolution of education and policy in the economy:

$$\begin{cases} e_{t+1}^i = \chi(e_t^i; \gamma_t, e_t^m) \\ \gamma_t = \Upsilon(e_t^m) \end{cases} \tag{3.24}$$

Two values solve the system above, and only one of them, $e_\infty = 0$, is found to be stable. The other long-run human capital level, below, is indeed unstable:

$$e_\infty = \left(\frac{\delta}{\Omega} - 1\right)\frac{e_\infty^m}{1-\kappa} \tag{3.25}$$

Given the convex shape of $\chi(\cdot)$ all the citizens with $e_0^i < e_\infty$ will converge, when $q < \bar{q}(\beta, \delta, \tau)$, to zero as quickly as e_∞^m, that in turn is a function of e_0^m, is small. The others will accumulate more and more and, as a result, the inequality dramatically increases.

The process can be attenuated in a sophisticated society. Here a high level of e_∞^m reduces the speed of convergence to $e_\infty = 0$, increasing the likelihood that some external shocks move the economy toward another development path. The model thus predicts multiple steady-states, one for naive societies that quickly converges to a poverty state, characterized by high inequality, thanks to populist politicians and one for sophisticated societies where political accountability impedes the advent of populism if the crises is not persistent enough.

However two societies with the same initial conditions, and in particular with the same initial distribution of education, can nonetheless be driven toward two different steady-states, depending on the level of optimism (q), the faith in the future (β), taxation (τ), inefficiency of the bureaucratic apparatus (κ), and human capital depreciation (δ). We summarize these results in Proposition 6. Proofs are gathered in the Appendix.

PROPOSITION 6: *When citizens are pessimists enough, i.e. if $q < \bar{q}(\beta, \delta, \tau)$, (or equivalently when the depreciation rate is above a critical value $\bar{\delta}(\beta, q, \tau)$), multiple stable steady-states arise, and initially less educated societies quickly converge to a state characterized by huge poverty, high inequality, and populist politicians. Conversely, if $q \geq \bar{q}(\beta, \delta, \tau)$ society get reacher and reacher, no matter initial conditions.*

3.4 Conclusions

The Chapter explores the rise and persistence of populist regimes in naive society that leaves room to selfish politicians to stay in power by promising that everything can be done and they can make them happen. This work particularly addresses the rise of the new wave of populism in Europe starting from the fall of the Berlin wall. The politics of Berlusconi and Bossi in Italy (followed up by the recent entry of Beppe Grillo with the Movimento 5 Stelle in politics), Haider in Austria and Le Pen in France have been mostly characterized by the paramount importance of comunication and promises that have capture the attention of the weakest part of the electorate. Although such promises embody all the charateristics of a cheap talking messages á la Crawford and Sobel (1982), the fortune of such politicians has been great indeed, especially Berlusconi's who reigned in Italy for about twenty years.

The explanation proposed by this Chapter relies on the ability of the electorate in keeping rulers accountable. Similarly to Chapter 2, political accountability pass through the political sophistication of the voters that defines the individual level of awareness over the type of the politician, and, in particular, on the pivotal role of the median voter. Once the median gets aware that such promises are unreliable he votes out the populist incumbent electing the challenger. In this case, we say the society is sophisticated enough to be ruled by a populist politician.

Despite the model follows a framework á la Ferejhon (1986), if the challenger would like to stay in power for a second appointment he must increase his willingness to tell the truth. Henceforth, despite the simultaneous equality between politicians, the model predicts a persistent disparity that increases over time and that in turn, by investment in human capital, favors a persistent process of sophistication, as we observe now the tail of it in the developed countries, or increasing inequality, as we rather observe in the Latin America Continent, depending on the initial distribution of the society in term of political sophistication. Thus the model additionally predicts disparity on the types of the politicians ruling between different countries.

The dynamics of the political/economic system, illustrated in sections , also sheds lights on the sustainability of the self-detrimental processes that have led Latin American poor to get poorer over time and eventually to loose everything as in Peronian Argentina and on the endogenous process that favor the advent of populism the more the society gets poor and naive. Although these processes apparently look as irrational, once we introduce tools of political psychology that puzzle brights itself: Naive voters are basically unaware of the politicians intentions providing to the latter opportunities for the manipulation of the economy and the electoral outcome. However, as far as sophisticated citizens are the majority, manipulation would be hard to be carried out.

According to that the model also proposes a new explanation for the rise of the inequality that not relies on the initial inequality itself, but on a political psychology process that potentially holds for any non degenerate initial distribution of education.

The understanding of the populism must pass through a better understanding of the society. There is no populism because there are no bad politicians in that country but because the society is sufficiently sophisticated to held a democracy. In other words, a mediocre political class is elected by a mediocre people in a democracy (Dahrendorf, 1996).

3.5 Proofs

3.5.1 Proof of Proposition 2

Claim 1: *there exists a unique optimal level of private investment as a function of the own level of education and the reckless rate of the ruler.*

Proof: The first order condition of maximization (3.11) is

$$(h_t^i)^{\phi-1} = \beta q V'(e_{t+1}^i)$$

that equals the costs of investing one unit more today, on the left hand side, to the expected marginal benefits from getting more educated in the future, on the right hand side: these are namely an increasing in future output and the catch up of higher political sophistication. The solution of the FOC uses the standard envelope condition to compute the expected future marginal benefits:

$$V'_{t+1} = \alpha(1-\tau)(q + (1-q)(1-\kappa)(1-\gamma_t)\eta_t^i)(e_{t+1}^i)^{\alpha-1}$$

Hence, we are left with

$$(h_t^i)^{\phi-1} = \alpha\beta(1-\tau)(q + (1-q)(1-\kappa)(1-\gamma_t)\eta_t^i)(e_{t+1}^i)^{\alpha-1}$$

To get an analytical solution we constrain the human capital elasticity $\alpha = 2$ and investment costs to be quadratic (i.e. $\phi = 2$). This yields equation (3.12). To prove that solution (3.12) is also unique we compute the second order condition ($\partial^2 V/\partial^2 h_t^i \leq 0$) that is satisfied iff

$$1 - \Gamma(q + (1-q)(1-\kappa)(1-\gamma_t)\eta_t^i) \geq 0$$

Since $(q + (1-q)(1-\kappa)(1-\gamma_t)\eta_t^i) \leq 1$, a sufficient condition for the SOC to hold is that $\Gamma \equiv 2\beta(1-\tau^H) \leq 1$.

Claim 2: *Citizens optimally respond to politicians' reckless rate by increasing private investments when $-\gamma_t$ increases.*

Proof: To prove that, we need to demonstrate that $\partial h_t^i/\partial\gamma_t \leq 0$. Let us define $A \equiv \Gamma(q + (1-q)(1-\kappa)(1-\gamma_t)\eta_t^i)$. Differentiation of (3.12) yields:

$$\begin{aligned}\frac{\partial h_t^i}{\partial \gamma_t} &= \frac{(1-\delta)e_t^i}{1-A}\frac{\partial A}{\partial \gamma_t} + \frac{(1-\delta)Ae_t^i}{(1-A)^2}\frac{\partial A}{\partial \gamma_t} \\ &= \frac{(1-\delta)e_t^i}{1-A}\frac{\partial A}{\partial \gamma_t}\Big(1+\frac{A}{1-A}\Big) \\ &\leq 0\end{aligned}$$

given that $\partial A/\partial \gamma_t \leq 0$.

Claim 3: *The reaction to a populist announcement is as strong as larger is the level of sophistication.*

Proof: The statement requires that $\partial^2 h_t^i/\partial \gamma_t \partial e_t^i \leq 0$. Straight differentiation yields:

$$\begin{aligned}\frac{\partial^2 h_t^i}{\partial \gamma_t \partial e_t^i} &= -\frac{(1-\delta)\Gamma(1-q)}{1-A}\Bigg[2\eta_t^i - \frac{1}{1-A}\Big[(1-\gamma_t)\frac{\partial A}{\partial \gamma_t}(1+3e_t^i) - 2\Omega e_t^i\Big] \\ &\quad - \frac{(1-\gamma_t)}{(1-A)^2}\frac{\partial A}{\partial \gamma_t}e_t^i\Big[\Omega - (1-\gamma_t)\frac{\partial A}{\partial \gamma_t}\Big]\Bigg] \\ &\leq 0\end{aligned}$$

given that $\partial A/\partial \gamma_t \leq 0$.

3.5.2 Proof of Proposition 3

First of all, we show that $e_{t+1}^i(e_t^i)$ is an increasing convex function. To see that note:

$$\frac{\partial e_{t+1}^i}{\partial e_t^i} = \frac{(1-\delta)(1-\Omega)}{\big[1-\Gamma(q+(1-q)(1-\kappa)(1-\gamma_t)\eta_t^i)\big]^2} \geq 0$$

$$\frac{\partial^2 e_{t+1}^i}{\partial^2 e_t^i} = \frac{(1-\delta)(1-\Omega)}{\big[1-\Gamma(q+(1-q)(1-\kappa)(1-\gamma_t)\eta_t^i)\big]^3}\frac{\Gamma(1-q)(1-\kappa)(1-\gamma_t)}{\bar{e}} \geq 0$$

where the second inequality holds by SOC, i.e. $\Gamma \leq 1$. Furthermore, it lays below the $45°$-line ($\partial e_{t+1}^i/\partial e_t^i \geq 1$) for $e_t^i \in (0, \bar{e}_t(\gamma_t))$, i.e.

$$e_t^i \geq \frac{(1-\Omega)^{1/2}[(1-\Omega)^{1/2}-(1-\delta)^{1/2}]}{\Gamma(1-q)(1-\kappa)(1-\gamma_t)}\bar{e} \equiv \bar{e}_t(\gamma_t).$$

Therefore by convexity we get, substituting in (3.13) $e^i_{t+1} = e^i_t = e_\infty$, two steady-state equilibria. If $\Omega \geq \delta$,

$$\begin{cases} e_\infty = 0 \qquad (stable) \\ e_\infty = \frac{\delta-\Omega}{\Gamma(1-q)(1-\kappa)(1-\gamma_\infty)}\bar{e} \geq 0 \qquad (unstable) \end{cases}$$

Otherwise,

$$\begin{cases} e_\infty = 0 \qquad (unstable) \\ e_\infty = \frac{\delta-\Omega}{\Gamma(1-q)(1-\kappa)(1-\gamma_\infty)}\bar{e} \leq 0 \qquad (stable) \end{cases}$$

Finally, taking the limit of e_∞ with respect to γ_∞ we get, in the most interesting case of $\Omega \leq \delta$:

$$\lim_{\gamma_\infty \to 0} e_\infty = \frac{\delta - \Omega}{(\Gamma - \Omega)(1-\kappa)}\bar{e}$$

$$\lim_{\gamma_\infty \to 1} e_\infty = \bar{e}$$

3.5.3 Proof of Proposition 4

Assume $\ln e^i_0 \sim \mathcal{N}(\mu_0, \Delta^2_0)$. To show that e^i_t keeps distributing lognormally over time let us call

$$\begin{aligned} \xi^i_0 &= 1 - \Gamma(q + (1-q)(1-\kappa)(1-\gamma_0)\eta^i_0) \\ &= (1-\Omega) + (\Gamma(1-q)(1-\kappa)(1-\gamma_0)/\bar{e})e^i_0 \end{aligned}$$

Since e^i_0 distribute lognormally, with mean μ_0 and variance Δ^2_0, and the other terms in ξ^i_0 are constants, ξ^i_0 must distribute lognormally as well with mean $(1-\Omega) - (\mu_0 + \ln\Gamma + \ln(1-q) + \ln(1-\kappa) + \ln(1-\gamma_0) - \ln\bar{e})$ and variance Δ^2_0. Finally, note that the sum of log-normal distributions yields, under broad conditions, another lognormal distribution.

If $\ln e^i_t \sim \mathcal{N}(\mu_t, \Delta^2_t)$ then $y^i_t = (e^i_t)^2$ is a transformation of e^i_t and must distribute as a lognormal too: $\ln y^i_t \sim \mathcal{N}(m_t, v^2_t)$, with $m_t = 2\mu_t + \Delta^2_t$.

The law of motion of aggregated level of human capital is obtained from (3.13). Taking a logarithmic transformation in both sides we are left with

$$\begin{aligned} \ln e^i_{t+1} &= \ln e^i_t + \ln(1-\delta) - \ln(1 - \Gamma(q + (1-q)(1-\kappa)(1-\gamma_t)\eta^i_t))) \\ &\approx \ln e^i_t - \delta + \Omega - \Gamma(1-q)(1-\kappa)(1-\gamma_t)\eta_t \end{aligned}$$

for small values of δ and $\Gamma(1-q)(1-\kappa)(1-\gamma_t)\eta_t$. Averaging yields:

$$\mu_{t+1} = \mu_t + \frac{\Delta_t^2}{2} + (\Omega - \delta) - \Gamma(1-q)(1-\kappa)(1-\gamma_t)\eta_t,$$

with $\eta_t \equiv \exp(\mu_t + \Delta_t^2/2)/\bar{e}$.

At the same way, we can compute the expected output in $t+1$ of the economy; taking a logarithmic transformation of (3.10) yields:

$$\begin{aligned}
\ln y_{t+1}^i &= 2\ln e_{t+1}^i + \ln\left[q\Big(1 + \frac{1-q}{q}(1-\kappa)(1-\gamma_t)\eta_t\Big)\right] \\
&\approx 2\ln e_{t+1}^i + \frac{1-q}{q}(1-\kappa)(1-\gamma_t)\eta_t + \ln(q)
\end{aligned}$$

Taking the average:

$$\begin{aligned}
m_{t+1} &= 2\mu_{t+1} + \frac{1-q}{q}(1-\kappa)(1-\gamma_t)\eta_t + \ln(q) \\
&= 2\mu_t + \Delta_t^2 + 2(\Omega-\delta) - 2\Gamma(1-q)(1-\kappa)(1-\gamma_t)\eta_t \\
&+ \frac{1-q}{q}(1-\kappa)(1-\gamma_t)\eta_t + \ln(q) \\
&= m_t + 2(\Omega-\delta) - 2\Gamma(1-q)(1-\kappa)(1-\gamma_t)\eta_t \\
&+ \frac{1-q}{q}(1-\kappa)(1-\gamma_t)\eta_t + \ln(q)
\end{aligned}$$

In time T2 the incumbent ruler faces the following maximization problem:

$$V_t^r(\gamma_t) = \max_{\gamma_t}\left\{\tau y_t - [q + (1-q)(1-\gamma_t)]b\tau y_t + \beta\mathbb{E}_t\big[V_{t+1}^r(\gamma_t)\big]\right\} \quad \text{s.t. (3.7)}$$

Taking a logarithmic transformation yields:

$$\begin{aligned}
\ln V_t^r(\gamma_t) &= \max_{\gamma_t}\Big\{\ln\tau + \ln y_t + \ln(1 - [q + (1-q)(1-\gamma_t)]b) \\
&+ \beta\mathbb{E}_t\big[\ln V_{t+1}^r(\gamma_t)\big]\Big\} \quad \text{s.t. (3.7)} \\
&\approx \max_{\gamma_t}\Big\{\ln\tau + \ln y_t - [q + (1-q)(1-\gamma_t)]b \\
&+ \beta\mathbb{E}_t\big[\ln V_{t+1}^r(\gamma_t)\big]\Big\} \quad \text{s.t. (3.7)} \\
&= \max_{\gamma_t}\Big\{\ln\tau + 2\mu_t + \Delta_t^2 - [q + (1-q)(1-\gamma_t)]b \\
&+ \beta\mathbb{E}_t\big[\ln V_{t+1}^r(\gamma_t)\big]\Big\} \quad \text{s.t. (3.7)}
\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}_t\big[\ln V^r_{t+1}(\gamma_t)\big] &= \ln q + \ln\tau + m_{t+1}\\ &= \ln q + \ln\tau + 2\mu_t + \Delta^2_t + 2(\Omega-\delta)\\ &+ (1-q)(1-\kappa)(1-\gamma_t)\eta_t\Big[2\Gamma - \frac{1}{q}\Big]\end{aligned}$$

Since the program is linear in γ_t, it is easy to note that

$$\frac{\partial \ln V^r_t}{\partial\gamma_t} = (1-q)b - \beta(1-q)(1-\kappa)\Big(\frac{1-2\Omega}{q}\Big)\eta_t$$

that is negative if and only if condition (3.20) is satisfied, that is if the cost of the public investment, relative to tax revenues, is small enough.

We now show that there exists a cutpoint-cost of the public good $\bar{b}$ such that for every $b > \bar{b}$ the ruler strictly prefer to going-for-broke instead of playing a strategy $\gamma_t = 1$. We first show that, at time $T2$, the value function of going-for-broke is strictly greater than the value taken by being responsible (note that, unlike the rent-seeking case, it is sufficient to show it for $\gamma_t = 1$ since the value function is increasing with γ_t for any $b > \underline{b}$). Let us call $\mathcal{V}^r_t \equiv \ln V^r_t(\gamma_t = 1)$.

$$\mathcal{V}^r_t < T_t$$

$$\ln\tau + 2\mu_t + \Delta^2_t > (1+\beta)\ln\tau + 2(1+\beta)\mu_t + (1+\beta)\Delta^2_t - qb + \beta\ln q + 2\beta(\Omega-\delta)$$

Solving by b yields $\bar{b}(\delta)$:

$$\bar{b}(\delta) \equiv \frac{1}{q}[\beta\ln\tau q + m_t + 2\beta(\Omega-\delta)].$$

Note that $\bar{b}(\delta)$ is a decreasing function of the depreciation rate of the human capital of the producers so as for high level of δ Go-for-Broke is more likely to be the optimal strategy in the stage game. In particular, there exists a threshold δ^* such that for any $\delta < \delta^*$ Going-for-Broke is not an admissible strategy for the ruler. To show that we need to solve the following inequality:

$$\bar{b}(\delta) > 1,$$

that holds for

$$\delta < \Omega + \frac{1}{2}\ln\tau q + \frac{1}{2\beta}(m_t - q) \equiv \delta^*.$$

3.5.4 Proof of Proposition 6

A stable steady state is a point $(e_\infty, \gamma_\infty)$ with the curve $\chi(\gamma)$ cuts the curve $\Upsilon(e)$ from above. An unstable steady state corresponds in each case to an intersection from below.

The dynamical system (3.23) reduces to a one-dimensional recursion: $e^i_{t+1} = \chi(e^i_t, \Upsilon(e^i_t))$. It has the following features:

(*i*)

$$\chi(0) = 0$$

(*ii*)

$$\chi'(e^i_t) = \frac{(1-\delta)(1-\Omega)}{\left[1-\Omega\Big(1+(1-\kappa)\frac{e^i_t}{e^m_t}\Big)\right]^2} \geq 0 \quad \text{by SOC}$$

(*iii*)

$$\chi''(e^i_t) = \frac{(1-\delta)(1-\Omega)(1-\kappa)}{\left[1-\Omega\Big(1+(1-\kappa)\frac{e^i_t}{e^m_t}\Big)\right]^3}\frac{\Omega}{e^m_t}$$

Given that SOC implies that the numerator is positive, we have that

$$\chi(e^i_t)\begin{cases}\text{is convex} & \text{if} \quad e^i_t < \Big(\frac{1}{\Omega}-1\Big)\frac{e^m_t}{1-\kappa}\\ \text{is concave} & \text{elsewhere}\end{cases}$$

The function $\chi(e^i_t)$ therefore presents an asymptotic behavior in correspondence of $e^i_t = (1/\Omega - 1)(e^m_t/(1-\kappa))$; on the left of such value the function diverges to infinity; on the right it diverges to $-\infty$. In this last region the function never crosses the horizontal axis nor the 45 degree line. Accordingly, we constraint the upper bound of the education support to be $\bar{e} < (1/\Omega - 1)(e^m_t/(1-\kappa))$, that amounts to exclusively consider the case in which Ω is small enough (that is the most interesting region for the rising of populist phenomena).

The steady states are rather computed as follow:

(*v*)

$$e^i_{t+1} = e^i_t = e_\infty$$

in four fixed points:

$$e_\infty^{(1)} = 0$$

$$e_\infty^{(2)} = \begin{cases} \left(\frac{\delta}{\Omega} - 1\right)\frac{e_\infty^m}{1-\kappa} \geq 0 & \text{iff} \quad \Omega \leq \delta \\ \left(\frac{\delta}{\Omega} - 1\right)\frac{e_\infty^m}{1-\kappa} \leq 0 & \text{elsewhere} \end{cases}$$

A fixed point e_∞ is stable if and only if

$$\left.\frac{\mathrm{d}\chi(e)}{\mathrm{d}e}\right|_{e=e_\infty} < 1.$$

Accordingly, it is straightforward to see that, when $\Omega \leq \delta$ (or if $q \leq \bar{q}(\beta, \delta, \tau) \equiv \frac{1}{2}\frac{\delta}{\beta(1-\tau)}$), $\chi(e_t^i)$ is a convex function that crosses the 45 degree line from below in correspondence of $e_\infty^{(2)}$ which is, in such a way, an unstable steady state, the unique stable one being $e_\infty^{(1)} = 0$. In the latter case ($\Omega(q) \geq \delta$), $e_\infty^{(2)} \leq 0$ is the unique stable fixed point, whereas, interestingly enough, $e_\infty^{(1)} = 0$ is unstable, meaning that each citizen with any initial condition $e_0^i > 0$ accumulate over time getting richer and richer.

In the former case ($\Omega(q) \leq \delta$), the most interesting in this chapter, rich citizens with $e_0^i > (\delta/\Omega - 1)(e_0^m/(1-\kappa))$ accumulate over time, eventually getting $\bar{e}$, whereas poor (and naive) citizens with $e_0^i < (\delta/\Omega - 1)(e_0^m/(1-\kappa))$ converge to a state of persistent poverty ($e_\infty^{(1)} = 0$). Note further that if

$$\left(\frac{\delta}{\Omega} - 1\right)\frac{e_0^m}{1-\kappa} \geq e_0^m$$

that is,

$$\delta \geq \Omega(2-\kappa) \iff q \leq \frac{1}{2}\frac{\delta}{\beta(1-\tau)(2-\kappa)}$$

poor are the majority and nonetheless they converge to a state of poverty and naivité by keeping support to the populist ruler.

$e_\infty^{(2)}$ depends on the trajectory of the median agent of the distribution, that is then pivotal. In time 1, from (3.23), is evident that $e_1^i \geq e_0^i$ if the citizen i is rich enough, i.e. if

$$e_0^i \geq \left(\frac{\delta}{\Omega} - 1\right)\frac{e_0^m}{1-\kappa}$$

by selecting only good investments to be done in $\sigma_t = H$ with responsible rulers. Naive agents instead will be driven by the psychology political process to carry out any investment including the bad investment proposed by populist rulers in state L. It leads eventually to a disastrous result with the poor getting poorer over time. This reasoning implies a divergence process according to which inequality broadens over time.

To see that note that the dynamics of the system turns into a complex inequality-spreading process described by the following individual linear law of motion:

$$e_t = \frac{(1-\delta)^t}{(1-\Omega)^t - (1-\Omega)^{t-1}\Omega e_0^i - \Omega(1-\delta)^{t-2}[1+\sum_{i=t-1}^{\infty}(t-i)(1-\Omega)]} e_0^m$$

which depends on the initial median level of distribution and on the initial individual level. This second dependence generates inequality in the long-run distribution.

Bibliography

[1] Acemoglu D., Egorov G., and Sonin K. (2013), *A Political Theory of Populism*, QJE, Vol. 128, No. 2: 771-805.

[2] Albertazzi D. and McDonnell D. (2008), *Twenty-First Century Populism: The Spectre of Western European Democracy*, New York and London: Palgrave Macmillan.

[3] Alesina A. (1988), *Credibility and Policy Convergence in a Two-Party System with Rational Voters*, AER, Vol. 78: 796-805.

[4] Banks, J. (1990), *A model of electoral competition with incomplete information*, JET, Vol. 50, No. 2: 309-325.

[5] Banks, J., Sundaram R. (1993), *Adverse Selection and Moral Hazard in a Repeated Elections Model*, in Political Economy: Institutions, Information, Competition, and Representation, eds. W. Barnett, M.J. Hinich and N. Schofield. New York, Cambridge University Press.

[6] Benabou R. (2013), *Groupthink: collective delusions in organizations and markets*, IZA Discussion Papers 7322, Institute for the Study of Labor.

[7] Benabou, R. and Tirole J. (1999), *Self-confidence: intrapersonal strategies*, IDEI mimeo.

[8] Benabou R., Tirole, J. (2002), *Self confidence and personal motivation*, QJE, Vol. 117, No. 3: 871-915.

[9] Canes-Wrone B., Herron M.C. and Shotts K.W. (2007), *Leadership and Pandering: A Theory of Executive Policymaking*, AJPS, Vol. 45, No. 3: 532-550.

[10] Connif M.L (2012), *Populism in Latin America*, University Alabama Press.

[11] Crawford V.P. and Sobel J. (1982), *Strategic Information Transmission*, Econometrica, Vol. 50, No. 6: 1431-1451.

[12] Dahrendorf R. (1996), *Introduction: Mediocre Elites Elected by Mediocre Peoples*, in: Elitism, populism, and European politics, ed. by Hayward J., Oxford : New York : Clarendon Press ; Oxford University Press.

[13] Dornbusch R. and Edwards S. (1991), *The Macroeconomics of Populism*, NBER Chapters, in: The Macroeconomics of Populism in Latin America, pages 7-13 National Bureau of Economic Research, Inc.

[14] Edwards S. (2010), *Left Behind: Latin America and the False Promise of Populism*, University Of Chicago Press.

[15] Fearon, J. (1999), *Electoral Accountability and the Control of Politicians: Selecting Good Candidates versus Sanctioning Poor Performance*, in Democracy, Accountability, and Representation, eds. A. Przeworski, S.C. Stokes and B. Manin, New York, Cambridge University Press.

[16] Ferejohn J. (1986), *Incumbent Performance and Electoral Control*, PC, Vol. 50, No. 1/3, Carnegie Papers on Political Economy, Volume 6: 5-25.

[17] Frisel L. (2009), *A theory of self-fulfilling political expectations*, JPE, Vol. 93, No. 56: 715-720.

[18] Glomm, G. and Ravikumar, B. (1992), *Public versus private investment in human capital endogenous growth and income inequality*, JPE, Vol. 100(4) :818-34.

[19] Gomez B.T. and Wilson J.M. (2001), *Political sophistication and economic voting in the American electorate: A theory of heterogeneous attribution*, AJPS: 899–914.

[20] Gradstein M. and Justman M. (1997), *Democratic choice of an education system: implications for growth and income distribution*, JEG, Vol. 2, No. 2: 169–183.

[21] Harrington, J.E. (1993), *Economic Policy, Economic Performance, and Elections*, AER, Vol. 83, No. 1: 27-42.

[22] Heidhues P. and Lagerlof J. (2003), *Hiding information in electoral competition*, GEB, Vol. 42, No. 1: 48-74.

[23] Hodler R., Loertscher S., Rohner D. (2003), *Inefficient policies and incumbency advantage*, JPE, Vol. 94, No. 9: 761-767.

[24] Knight A. (1998), *Populism and Neo-populism in Latin America, especially Mexico*, Journal of Latin American Studies, Vol. 30, No. 2: 223-248.

[25] Levitsky S. and Roberts K.M. (2011), *The Resurgence of the Latin American Left*, Johns Hopkins University Press.

[26] Maskin E. and Tirole J. (2004), *The Politician and the Judge: Accountability in Government*, AER, Vol. 94(4): 1034-1054.

[27] Panizza F. (2005), *Introduction: populism and the mirror of democracy*, in: Panizza, Francesco, (ed.) Populism and the Mirror of Democracy, Phronesis. Verso Books, London, UK, 1-31

Chapter 4

The Italian primary school-size distribution and the city-size: a complex nexus

4.1 Introduction

There is a growing literature that nowadays sheds light on complexity features of social systems. Notable examples are firms and cities [1, 2, 3, 4], but many others have been proposed [5, 6]. These systems are perpetually out of balance, where anything can happen within well-defined statistical laws [7, 8]. Italian schools system seems to not escape from the same characterization and destiny. Despite several attempts of the Italian Ministry of education to reduce the class-size to comply with requirements stated by law [9, 10, 11], no improvements have been made and still heterogeneity naturally keeps featuring the size distribution of the Italian primary schools.

In this chapter we characterize the statistical law according to which the size of the Italian primary schools distributes. Using a database provided by the Italian Ministry of education in 2010 we show that the Italian primary school-size approximately distributes (in terms of students) as a log-normal distribution, with a fat lower tail that collects a large number of very small schools. Similarly to the firm-size [13], we also find the upper tail to decrease exponentially. Moreover, the distribution of the school growth rates are distributed with a Laplassian PDF. These distributions are consistent with the Bose-Einstein preferential attachment process. These results are found both at a provincial level and aggregate up to a national level, i.e. they are universal and do not depend on the geographic area.

The body of the distribution features a bimodal shape suggesting some source of heterogeneity in the school organization. The evidence of the bimodality underlies the interplay between different processes that define thresholds and boundaries that are very peculiar for the Italian primary school-size distribution. The question that we attempt to address in this chapter is whether such regularity might depend on the complex geographic features of the Country that in turn determines the way population (and in particular young people) distributes. We address these questions by analyzing in depth the spatial distribution of the schools, with particular regard to the areas where commuting is more effortful. We then proceed by investigating the complex link between schools and comuni, the smallest administrative centers in Italy, addressed by the introduction of a new binning methodology and a new spatial interaction analysis. Our conclusions indicate that the bimodality of the Italian primary school-size distribution is very likely to be due to a mixture of two laws governing small schools in the countryside and bigger ones in the cities, respectively.

Several examples of different regional schooling organizations are analyzed and discussed. We use GPS code positions for schools in two very different Italian Regions: Abruzzo and Tuscany. We introduce a measure of the average spatial interaction intensity between a school and the surrounding ones. We show that in regions like Abruzzo, that are mainly countryside, a policy favoring small schools uniformly distributed across small comuni has been implemented. Abruzzo small schools are generally located in low density populated zones, in correspondence of very small comuni. They are also very likely to have another small school as closest and the median distance between them is 8 km that is also the distance between small comuni. In Tuscany, a flatter region with a very densely populated zone along the metropolitan area composed by Florence, Pisa and Livorno, we conversely find:

- higher school density;
- stronger interaction between small and big schools;
- greater average proximity among schools.

We address these stylized facts by arguing that the Italian primary school organization is basically the result of a random process in the school choice made by the parents. Primary education is not felt so much determinant to drive housing choice, like in US, because of the absence of any territorial constraint in school choice. Even if there is a certain mobility *within* a comune toward the most appealing schools, primary students generally do not move *across* comuni to attend a school. As a result, school density and school-size are prevalently driven by the population density and then by the geographical features of the territory. This generates a mixture in the schooling organization that turns into a bimodal shape distribution.

4.2 Results

Empirical evidence

We analyze a database on the primary school-size distribution in Italy that provides information on public and private schools, locations, and the number of classes and students enrolled. Data are collected, at the beginning of every academic year, by the Italian Ministry of education to be used for official notices. Our dataset covers $N = 17187$ primary schools in 2010 of which 91.31% were public. Almost seven thousands are located in mountain territories, (which represent the 40%) and 4101 are spread among administrative centers (provincial head-towns).

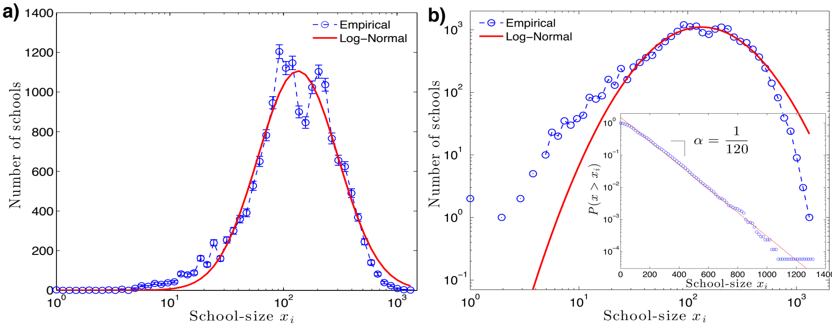


Figure 4.1: School-size distribution. a. Italian primary school-size distribution according to the number x_i of student per school $i \in [1, \ldots, N]$ for the year 2010. The empirical distribution is drawn in blue (each circle is a bin); the red line stands for the Gaussian fit with mean $\hat{\mu} = 4.77$ ($\hat{\mu}/\ln(10) = 2.07$) and standard deviation $\hat{\sigma} = 0.85$ ($\hat{\sigma}/\ln(10) = 0.37$). On a non-logarithmic scale, $\exp(\hat{\mu}) = 118$ and $\exp(\hat{\sigma}) = 2.34$. $N = 17187$. Statistical errors (SE) are drawn in correspondence of each bin, according to $\sqrt{N_{bin}}$. SE are bigger in the body of the distribution and tinier in the tails. Nevertheless, central bins space from the two peaks, $m_1 = 1.7$ and $m_2 = 2.3$, at least 6 times the SE, equals on average to $\sqrt{10^3} = 32$. In this case the probability to have a non bimodal shape under our distribution is 4×10^{-15}. **b.** Italian primary school-size distribution in log-log scale. As expected, the theoretical distribution has drawn as a perfect parabola (the red curve), $y = ax^2 + bx + c$, such that $\hat{\mu} = -b/2a$ and $\hat{\sigma} = -1/2a$. Conversely, the empirical distribution does not plot as a parabola, at least for what regards to the tails which deviate from the log-normal. The inset figure shows a functional form of the right tail of the empirical distribution. We plot the cumulative distribution, $P(X > x_i) = \exp(-\alpha x_i)$, of school sizes in semi-logarithmic scale with characteristics size $\alpha = 0.0084$. This in turn means that there are approximately 120 students per school.

In Italy primary education is compulsory for children aged from six to ten. However, the parents are allowed to choose any school which they prefer, not necessarily the school closest to their home, [14]. We define x_i the size of the school $i \in [1, \ldots, N]$ as the number of students enrolled in each school. Fig. 4.1(a) shows the histogram of the logarithm of the size of all primary schools in Italy. The red solid curve is the log-normal fit to the data

$$P(\ln x) = \exp\left(-\frac{(\ln x - \hat{\mu})^2}{2\hat{\sigma}^2}\right)\frac{1}{\sqrt{2\pi}\hat{\sigma}} \tag{4.1}$$

using the estimated parameters $\hat{\mu} = 4.77$ ($\hat{\mu}/\ln(10) = 2.07$), the mean of the $\ln x$ of the number of students per school, and its standard deviation, $\hat{\sigma} = 0.85$ ($\hat{\sigma}/\ln(10) = 0.37$). On a non-logarithmic scale, $\exp(\hat{\mu}) = 118$ and $\exp(\hat{\sigma}) = 2.34$ are called the location parameter and the scale parameter, respectively [15]. The histogram in Fig. 4.1a suggests that log-normal fits data quite well. However, even a quick glance reveals that there are too many schools with a small dimension and much less mass in the upper tail with respect to the fit, suggesting that the number of students of the largest schools is smaller than would be the case for a true log-normal. In other words, similarly with firms-size distribution [16], tails seem to distribute differently from the log-normal distribution. Also Fig. 4.1a reveals a bimodal shape of the school-size distribution that we will extensively investigate below.

These findings can be detected in a more powerful way by plotting the histogram in a double logarithmic scale, comparing the tails of the log-normal distribution with those of the empirical one. We do this in Fig. 4.1b where y-axes represents the logarithm of the number of schools in the bins whereas in the x-axes the logarithm of the number of students stands. The empirical distribution differs significantly from the theoretical distribution which is a perfect parabola (the red curve), both in the tails and in the central bimodal part. A functional form of the right tail of the empirical distribution is revealed in the inset of Fig. 4.1b where we plot the cumulative distribution $P(X > x)$ of school sizes in semi-logarithmic scale. The straight line fit suggests that the right tail decreases exponentially $P(X > x) = \exp(-x\alpha)$ with a characteristics size $\alpha = \frac{1}{120}$. This in turn means that there are approximately 120 students per school and also that the distribution of large schools declines exponentially. The exponential decay of the right tail of size distribution is consistent with Bose-Einstein preferential attachment process and is observed in the distribution of sizes of universities and firms.

Next we investigate the growth rates of elementary schools. Since temporal data are not currently available, we look at the single academic year, the 2010, and define the growth rate g_i as follows:

$$g_i \equiv \frac{x_i^1 - x_i^5}{\sum_{j=1}^{5} x_i^j} = \lambda_i - \mu_i, \tag{4.2}$$

where x_i^j stands for the number of students attending the *j-th* grade in school i, with $j \in [1, 5]$; $\lambda_i \equiv x_i^1 / \sum_{j=1}^{5} x_i^j$ is the fraction of students that have been enrolled in the first grade at six years old in school i, whereas $\mu_i \equiv x_i^5 / \sum_{j=1}^{5} x_i^j$ is the fraction of students that exit the school after the 5-*th* grade. Fig. 4.2a shows the relation between growth rate g_i and school-size x_i. The numbers of grades j provided by each school i, named J_i, is defined by the color gradient bar on the right side of the Fig. 4.2a. Blue circles identify schools with $J_i = 1$. Such a group collects schools just established only providing the 1-*st* grade, i.e. with $\lambda_i = 1$ *and* $\mu_i = 0$, or that are going to close providing only the 5-*th* grade, i.e. with $\mu_i = 1$ *and* $\lambda_i = 0$. As soon as more grades are provided (colors switching to the warm side of the bar) schools tend to cluster around a null growth rate.

In Fig. 4.2b we investigate the growth/size relationship in depth. We demonstrate the applicability of the Gibrat law that states that the average growth rate is independent on the size [17, 18]. We define the average of the school size in each bin c as $\langle x_i \rangle_c$. The number of school in each bin n_c is represented by the size of the circle and the average number of grades $\langle J_i \rangle_c$ is depicted according to the color gradient on the right side (the same of Fig. 4.2a). Independently from the size and the number of grades provided, schools do not grow on average. Nevertheless, we find more variability in smaller schools, apart from schools with $x_i < 10$, namely hospital-based schools mostly similar to one another, and the standard deviation of the growth rate $\sigma_{g(\langle x_i \rangle_c)}$ is found to be decreasing as $\langle x_i \rangle_c^{-\beta}$ with school-size by a rate of $\beta \approx .60$ (subFig. 4.2b inset). This is consistent with what has been found for other complex systems like firms or cities [13, 20, 21, 22, 23].

In Fig. 4.3a we study the growth rate distribution, where the probability density function $P(g = g_i)$ of growth rate has been plotted. The blue line represents the full sample (all the schools) distribution. Black

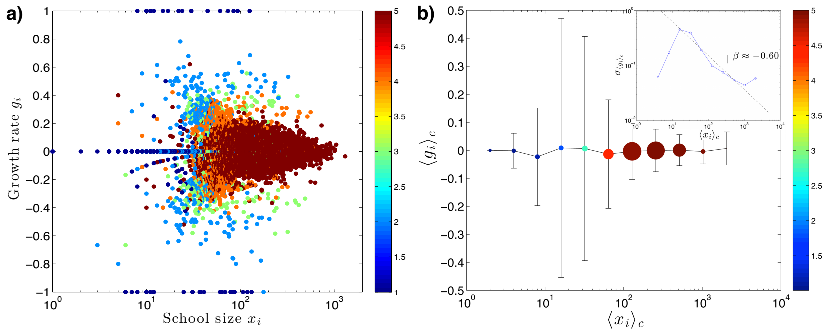


Figure 4.2: The growth rate and school-size relationship. The growth rate g_i is defined according to Eq. (4.2). **a.** Colors, according to the vertical bar on the right-hand side of the graph, are the number of grades J_i provided by the school i. Smaller schools (in blue) with $J_i = 1$ are both the newest one (just created, with $\lambda = 1$) and schools that are going to close (with $\mu = 1$). They can also be schools that do not grow yet providing just one grade (i.e. $j = 3$). **b.** The mean growth rate clusters around zero across different subsets c that are differently populated by n_c schools according to the size of the circles. The color of the circles stand for the average number of grades J_i (the same gradient color bar of Fig. 2(a) is used here). The variability within each cluster c is shown in the inset figure. Apart from schools with $x_i < 10$, namely hospital-based schools mostly similar to one another, the standard deviation is found to be decreasing with school-size by a rate of $\beta \approx .60$.

and red colors identify the full capacity schools ($J_i = 5$) and the schools with $J_i < 5$, respectively. Regardless of the number of grades provided, the growth distribution underlines a Laplace PDF in the central part of the sample [24]. The not-fully covered schools show a three peak behavior, where the left peak represents schools which are going to close, the central peak gathers schools that provide several grades but still in equilibrium phase, and the right peak is made up by the growing schools. Fig. 4.3b reports empirical tests for the tails of the PDF of the growth rate of the full sample (the upper one in blue, and the lower one in black). The asymptotic behavior of g can be well approximated by power laws with exponents $\zeta \approx 4$ (the magenta dashed line), bringing support to the hypothesis of a stable dynamics of the process [20]. All these findings are consistent with the Bose-Einstein process according to which the size distribution has an exponential right tail, a tent-shaped distributed growth

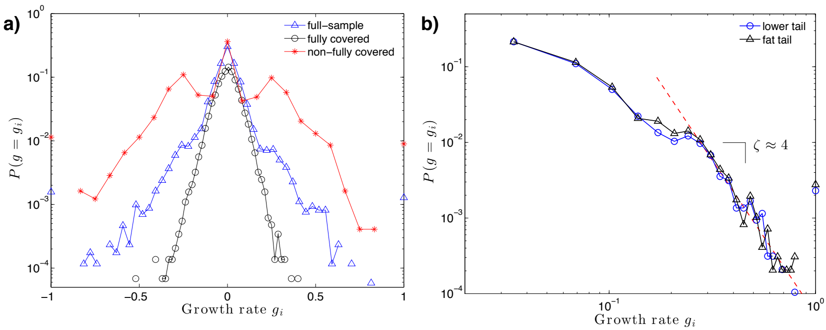


Figure 4.3: The growth rate distribution of the Italian primary schools in 2010. a. The probability density function $P(g = g_i)$ of growth rate has been plotted underlying a Laplace PDF in the body around $P(g) = 1$ and $P(g) \approx 10^{-1.5}$. Blue triangles ($\triangle$) stand for the full sample distribution, black circles ($\circ$) indicate mature schools with $J_i = 5$, and red stars ($*$) schools with $J_i = 1$. **b.** The plot reports empirical tests for the tails parts of the PDF of growth rate, the upper one in blue ($\circ$), and the lower one in black ($\Box$). The asymptotic behavior of g can be well approximated by power laws with exponents $\zeta \approx 4$ (the magenta dashed line).

rate g_i, with a Laplace cap and power law tails, the average growth rate is independent of the size, and the size-variance relationship is governed by the power law behavior with exponent $\beta \approx 0.5$ [25].

4.2.1 City size and school size

Fig. 4.1a features the coexistence of two peaks, the first peak corresponding to $\log_{10} x_i \equiv m_1 = 1.7$ and the second one to $\log_{10} x_i \equiv m_2 = 2.3$, divided by a splitting point in correspondence of $\log_{10} x_i \equiv \bar{m} \approx 2.1$. The school sizes corresponding to these features are $\mu_1 = 10^{m_1} = 50$, $\mu_2 = 10^{m_2} = 200$, and $\bar{\mu} = 10^{\bar{m}} = 128$, with $\bar{\mu}$ approximately equal to the average school size. 39% of the Italian primary schools distribute on the right of $\bar{\mu}$, and more than 60% distribute on the left side. We test the alternative hypothesis of unimodality by looking at the probability that the numbers of schools in the two central bins n_1, n_2 are not smaller and the numbers of schools in the next three bins n_3, n_4, n_5 are not larger than a certain number n^* provided that the standard deviation of the number of schools in these bins due to small statistics is $\sqrt{n^*}$. This probability is equal to $p(n^*) = \prod_i \mathrm{erfc}(|n_i - n^*|/\sqrt{2n^*})/2$ and it reaches maximum $p_{\max} \approx 4 \times 10^{-15}$ at $n^* = 980$. Accordingly, we establish the bimodality with a very high confidence. This is also consistent with the bimodality index that we find to be equal to $\delta = (\mu_1 - \mu_2)/\sigma = .45$, [26].

In this section we investigate the source of this heterogeneity that we find to be related to geographical and political features of the country and remarkably to the size of the comuni, the smallest administrative centers in Italy (information on comuni are provided by the Italian statistical institute, ISTAT), also here referred interchangeably as cities regardless of the size, p_k. A particular treatment is devoted to the nexus between the school-type (private versus public) and the geographical features of the comuni in the supplementary information, where we show that private schools are much less variable in size than public schools and have a narrow unimodal distribution peaked at approximately 100 students which contributes to the left peak of the entire school size distribution (Figure 4.14).

We denote a comune with letter $k = [1, \ldots, K]$. In 2010, $K = 8,092$ comuni have been counted in Italy, the 40% of which located in the mountains. We define $\mathcal{M}$ the set of mountains comuni and, accordingly, we call school i a mountain school iff it resides in a comune $k \in \mathcal{M}$ (in the sec. 4.5 we explain the mechanism according to which Italian comuni are classified as mountains). Each city k has $n_k \geq 0$ schools (more than 15% of the cities have no schools) and population p_k, which distributes approximately as a log-normal PDF (see Fig. 4.4a), except for the right tail that is distributed according to a Zipf law, i.e. $p_k \sim r(p_k)^{-\xi}$ with slope $\xi \approx 1$ [2, 3, 27, 28, 29]. In Fig. 4.4b we find $\xi \approx .80$, in Italy, that is exactly the slope of the power law $p_k \sim r(n_k)^{-\zeta}$ which links the population p_k with the rank of this city in terms of number of schools n_k (blue circles in Fig. 4.4b), i.e. $\zeta = \xi \approx .80$. This means that the first city, Rome, has almost the double number of schools than Milan, and triple of Naples, while Rome has almost the double of inhabitants of Milan, and the triple of Naples. This amounts to say that n_k is a good proxy for the city-size.

We use the number of schools to assign comuni to different clusters

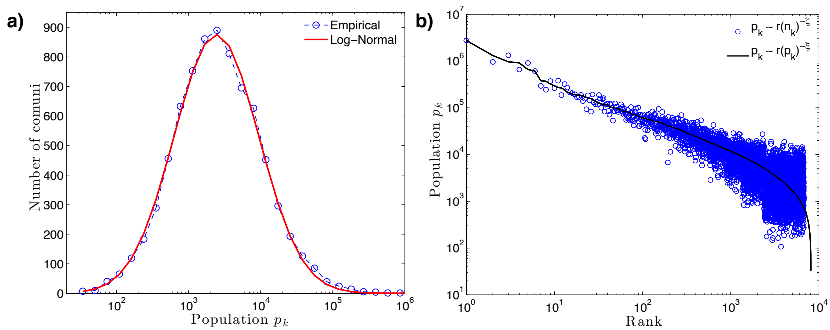


Figure 4.4: Cities features. a. The Italian city-size distribution for $K = 8092$ observations. Blue circles stand for each city-bin whereas the red solid line draws the log-normal fit of the data. Conversely to the school-size distribution depicted in Fig. 4.1a, the city-size PDF features single-peakedness, but similarly it has a power-law decay in the upper tail. **b.** Zipf plot for Italian cities according to the size p_k and the number of schools n_k. The black line draws the classical Zipf plot $p_k \sim r(p_k)^{-\xi}$, with cities ranked according to population p_k. Blue circles instead depict the Zipf plot $p_k \sim r(n_k)^{-\zeta}$, with cities ranked according to the number of schools n_k. Consequently, the sample reduces to $M = 6726$ over $N = 8092$ since more of the 15% of the cities have no schools.

$h \in [1, \ldots, H]$, according to

$$h = \{\forall\, k \in [1, \ldots, K] : 2^{h-1} \leq n_k < 2^h\}. \tag{4.3}$$

Accordingly, the first bin $h = 1$ gathers all the comuni with only one school; the second one collects all the comuni with $n_k = [2, 3]$, and so on. Though we find the average population $\langle p \rangle_h$ to increase across different city-clusters h, less comuni K_h lie in more populated clusters (the magenta and black lines in Fig. 4.5a). Interestingly, we find the interaction term $K_h \langle p \rangle_h$, the green line in Fig. 4.5a, to distribute uniformly across different comuni-clusters, meaning that in small comuni with $n_k = 1$ live the same population than in bigger ones with much more schools.

Nevertheless, population is differently composed across city-clusters and a smaller fraction of young people is found in smaller comuni. To see that we also introduce a clusterization of comuni according to population. Each comune is assigned to a cluster $c \in [1, \ldots, C]$ composed by

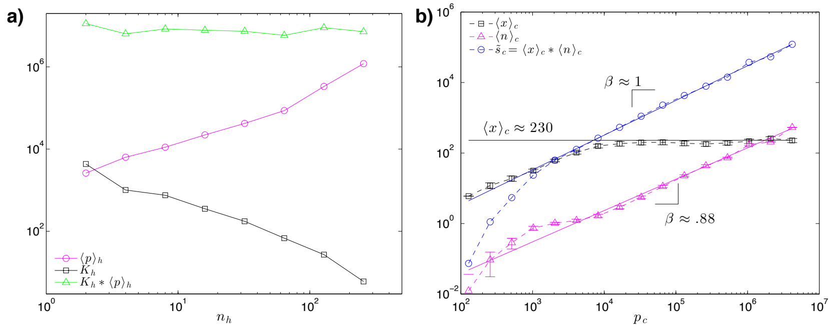


Figure 4.5: Population features. a. Each comune is assigned to 8 clusters, according to Eq. 4.3, and scattered against population, the magenta line (○) and the number of cities K_h, the black line (□). The interaction term, $K_h * \langle p \rangle_h$, the green line (△), represents the total population living in each city-cluster h. **b.** According to Eq. 4.4 K cities are assigned to $C = 16$ clusters. In the x-axis the number of inhabitants in cluster $c = \{7, 22\}$ is scattered against the average number of schools (magenta line (△)) and the average school-size $\langle x \rangle_c$ (the black line (□)). The interaction term (○), representing the typical number of schooling-aged population in cluster c, $\tilde{s_c} = \langle x \rangle_c * \langle n \rangle_c$ distributes as a power law with coefficient $\beta \approx 1$ for cities bigger than 10^3 inhabitants, and it is drawn in green. For smaller comuni, instead, the line drops meaning that a smaller fraction of young people features them.

all the comuni k with population p_k ranging from ψ^{c-1} to ψ^c, i.e.

$$c = \{\forall\, k \in [1, \ldots, K] :\ \psi^{c-1} < p_k \le \psi^c\}. \tag{4.4}$$

Setting the parameter[1] $\psi = 2$ yields $C = 23$ clusters. Although the first seven sets are empty because no comuni in Italy has less than 128 inhabitants, the first (non-empty) cluster, $c = 8$, collects very small comuni with $p_k \in (128, 256]$. The last one, $c = 23$, conversely, is composed by the biggest cities with $p_k \in (2^{22}, 2^{23}]$. In Fig. 4.5b we plot the average number of schools $\langle n \rangle_c$ (magenta line) and the average school-size $\langle x \rangle_c$ (the blue line) against the comuni size p_c for each non-empty cluster c. We find that the average number of schools increases as a power law with coefficient $\beta = 0.88$. This is consistent with the literature [2, 3, 28, 29] that

[1] It is possible to change the value of ψ without having any effect on the shape of the

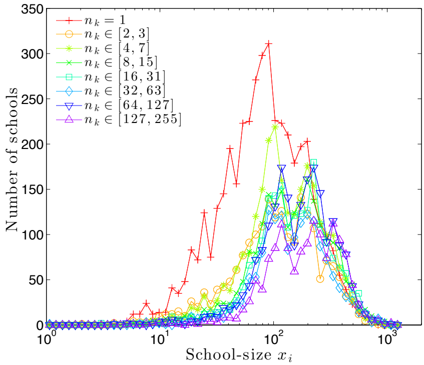


Figure 4.6: City vs. number schools. School-size distribution for different city-samples clustered according to the number of schools, i.e. to Eq. 4.3. Only comuni with $n_k = 1$ show a single peak school-size distribution, clustered around m_1 (the $+$-red line on the top). They have an average population of 2000 inhabitants and the 81% are located in mountain territories.

has stressed the emergence of scale-invariant laws that characterize the city-size distribution. The average school-size increases with the population of the city reaching an asymptotic value at $\langle x \rangle_c \simeq 230$ students per school in the large cities. As expected, the interaction term, representing the average number of school-aged population in comuni belonging to cluster c, $\tilde{s}_c = \langle x \rangle_c * \langle n \rangle_c$, behaves linearly with the comuni size except for small comuni with $p_c < 10^3$, for which the school-aged population constitutes a smaller fraction of the total population than in large cities.

In Fig. 4.6-4.7 we investigate the school-size distribution according to the comuni features. To this end, Fig. 4.6 draws the distributions of $\log_{10} x_i$ conditionally on the number of schools, n_k, in the comune k. It yields 8 curves, one for each cluster h defined in Eq. 4.3. The first cluster is drawn in red ($+$) distributing all the schools located in comuni where only one school is provided. The orange line ($\circ$) distributes all the schools provided in comuni with two or three schools (i.e. $h = 2$); and so on. The interesting point of Fig. 4.6 is that only the school-size distribution of the smallest comuni (with $n_k = 1$) features a unimodal

---

distributions

shape. The reason for that relies on the fact that comuni with only one school are geographically similar: they are the 57% of the total, with little more than 2000 inhabitants, the 81% of which are located in mountain territories.

The relationship between school-size and altitude is investigated in Fig. 4.7a. Instead of conditioning on $\mathcal{M}$, here we propose a more consistent exercise according to which comuni are assigned to different bins on the basis of the altitude. In such a way, we can analyze comuni with 1,000 meters above the sea differently to those with 600 meters of altitude that would be gathered in the same cluster $\mathcal{M}$ throwing away informative heterogeneity. It yields 5 bins: the first bin (drawn as a blue $+$ line) gathers all the comuni whose altitude is lower than 125 meters above the see level (labeled 125 in Fig. 4.7a). Comuni with an altitude between 125 and 250 meters above the sea level composed the second bin (the green $\circ$ line). These two distributions cluster around the second mode m_2 and in the Supplementary Information we additionally demonstrate that the hypothesis of bimodality can be rejected for the latter distribution. However, the greater the altitude of the comuni the greater is the shift of the correspnding school-size distribution toward the small schools and the greater is the contribution of these comuni to the first mode m_1. Such a shift becomes evident for comuni with an altitude between $250m$ and $500m$ (red $\triangle$ line). Comuni located between $500m$ and $1000m$ (cyan $\times$ line) and above $1000m$ (purple $\square$ line) clusterize around m_1.

Even the largest cities are very different from each other in terms of their school size distribution. This heterogeneity is very likely to be driven by geographical features. where we restrict our interest on the largest Italian cities belonging to cluster $h = 8$ (and to the first two bins in terms of altitude in Fig. 4.7a). These cities provide a number of schools n_k between 127 and 255, whose overall size distribution shows a three-peak shape with a third peak around 300 students absent in smaller cities (the bottom violet $\triangle$ line in Fig. 4.6). The presence of the three peaks around 100, 200 and 300 students might suggest the presence of architectural standards of school buildings supporting these particular sizes. However, by plotting the distribution by city, Fig. 4.7b, we show that

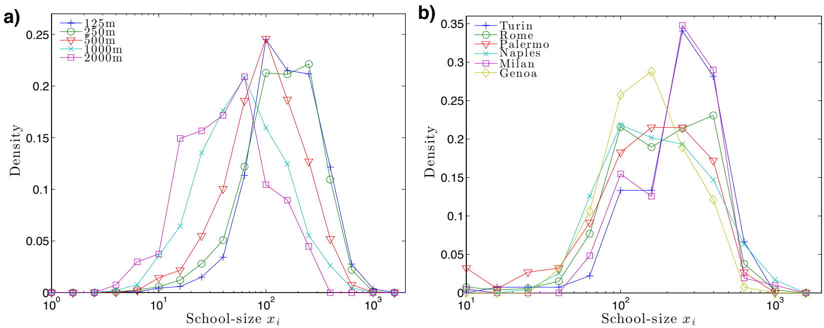


Figure 4.7: School-size distribution conditional on comuni features. a. School-size distribution for different city-samples clustered according to the altitude. The altitude of the comune shift the school-size distribution (shift location effect) as higher comuni are generally smaller schools. **b.** School-size distribution in the six biggest Italian cities. With the exception of Rome, the hypothesis of unimodality may not be rejected in none of the biggest cities. In particular, flatter cities, such as Milano and Torino, mostly contribute to second mode m_2, whereas in Genova, Italian city built upon mountains that steeply ended on the sea, all the school-size distribution stands on the left side.

all the traces of trimodality disappear. In particular flatter cities, such as Milano and Torino, mostly contribute to second mode m_2, whereas in Genova, an Italian city built upon mountains that steeply slope towards the sea, the school-size distribution is unimodal contributing mostly to the first mode m_1.

Another way to look at the effect of geography on the comunal school-size is to compute the fraction of large schools on the total within each comune k:

$$P_k(x_i > \bar{\mu} | \forall\, i \in k) \equiv \frac{n_k(x_i > \bar{\mu})}{n_k} \quad \forall\, i \in k, \tag{4.5}$$

where $n_k(x_i > \bar{\mu})$ stands for the number of schools that, in each comune k, are larger than the minimum $\bar{\mu}$ of the school-size distribution shown in Fig. 4.1a. It can also be interpreted as the contribution rate of a comune k to the second mode m_2. The upper panel of Fig. 4.8 diagrammatically explains how $P_k(\cdot)$ is computed.

We firstly study the relationship between $P_k(\cdot)$ and population, then looking at the spatial distribution across the Italy. In Fig. 4.8, we clus-

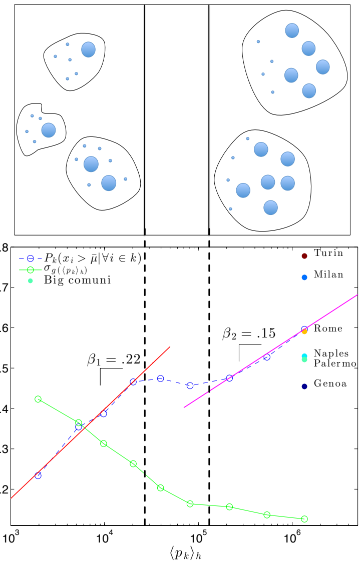


Figure 4.8: Fraction of large schools in comune k. The panel above shows the process according to which each comune, with population p_k defined by the size of the the black circles, is assigned to either patterns on the basis of the size of the schools provided in there (the small blue circles). The panel below shows that more populated clusters of cities are, on average, more likely to have schools sized around m_2. The relationship, depicted in blue, is however non monotonic. In correspondence of each bin h, the standard deviations has been computed, underlining the outstanding variability in very small cities (the green line).

terize comuni according to Eq. 4.3, and for each bin h we compute the average $\langle P_k(x_i > \bar{\mu} | \forall i \in k) \rangle_h$ and population $\langle p_k \rangle_h$. Interestingly, the plot shows that $P_k(\cdot)$ does not increase monotonically with population, demonstrating the existence of two city-patterns. More precisely, cities with less than 10^4 inhabitants follow a pattern according to which the fraction of big schools, with $x_i > \bar{\mu}$, increases, on average, with population at a rate of $\beta_1 \approx .22$; in cities with more than 10^5 we find the effect of population to be smaller, corresponding to $\beta_2 \approx .15$. For the cities with population between 10^4 and 10^5, the fraction of large schools does not increase with size suggesting that exogenous shocks such as altitude,

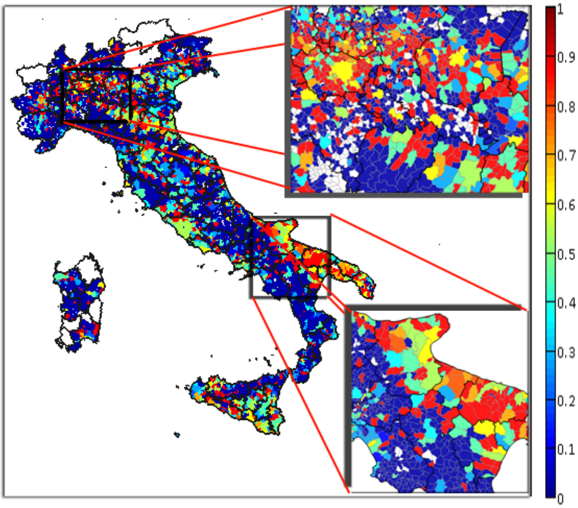


Figure 4.9: Spatial distribution of cities according to $P_k(x_i > \bar{\mu}|\forall\, i \in k)$. Warmer territories stand for cities more likely of having schools distributed around m_2. The two figure inset underline the region around Milan (in the North), on the top, and the regions of Basilicata (mostly mountain, at the left side) and of Apulia (mostly flat, at the right side), on the bottom. Maps generated with Matlab.

rugged terrain and age might shift a city in this transition zone to either mode m_1 or mode m_2.

Overall, the distribution of $P_k(x_i > \bar{\mu}|\forall\, i \in k)$ is strongly correlated with the geographical features of the comuni territory. The map in Fig. 4.9 clarifies this point; all the mountain territories, Apennines that represent the spine of the peninsula and the Alps on the northern side, turns to be comuni with small schools, since the share of small schools in mountain comuni is equal to $P(x_i \leq \bar{\mu}|k \in \mathcal{M}) = 0.72$. As soon as the probability to contribute to m_2 increases the colors get warmer; but this is very unlikely to be in mountain territories, because less than 30% of mountain comuni contribute to the antimode. Some regional patterns are also shown in the insets. The first upper panel depicts the area around Milan, which is surrounded by warm colors that mostly dye the Pianura Padana around. On the south side, Apennines approach and colors get blue with a lot of comuni with no schools (depicted in white). This pattern is more evident in the lower panel, which maps the region of Apulia, flat and mostly red, and the Basilicata on the left side, mountainous and mostly blue colored.

4.2.2 Countryside versus dense regions

In this last section, we bring more evidence on the effect of geography and comuni organization on the school-size by restricting our attention at two Italian regions: Abruzzo and Tuscany. But same results stand by looking at regions with the same geographical features. The two regions have very peculiar and representative geographical and administrative characteristics. Abruzzo is a mostly mountain region with a little flat seaside; it has four main head towns divided from each other by mountains. Conversely, Tuscany has many flat zones in the center and the mountain areas shape the region boundaries. Remarkably, it has a very high densely populated zone along the metropolitan area composed by Florence, Pisa and Livorno.

These two regions also differ in terms of administrative organizations, Abruzzo favoring the establishment of comuni with a smaller size due to the presence of mountains. Fig. 4.10 shows the comuni population distributions in Tuscany and Abruzzo. We clusterize comuni using the algorithm in Eq. 4.4. As Fig. 4.10a makes clear, comuni distribute approximately as a log-normal pdf in both regions, i.e. as a parabola in a log-log scale (the green-$\circ$ line stands for Abruzzo pdf, the magenta-$\triangle$ for Tuscany). Nevertheless, Tuscany has bigger cities. Figure 4.10a also shows the average number of schools as function of the population. The fact that $\langle n_K \rangle$ is less than one for small comuni reflects the fact that many of these comuni do not provide schools. Abruzzo has a larger number of small comuni that do not provide schools. The first bin collects comuni with a bit more than 100 inhabitants. They are 7 in Abruzzo (none in Tuscany), none of them providing any school services. The second bin accumulates 10 comuni in Abruzzo with 300 inhabitants (none in Tuscany), of which only one has a school. Comuni with about 600 inhabitants are 40

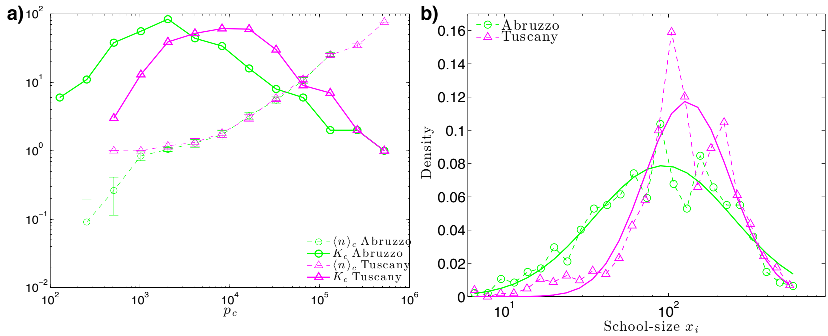


Figure 4.10: Regional analysis I. a. The figure distributes the city-size in Abruzzo (○-green) and Tuscany (△-magenta) by plotting the number of comuni, K_c, against the number of inhabitants, p_c. Also shown is the average number of schools in a comune in Abruzzo and Tuscany, belonging to a bin c defined by Eq. 4.4, by the circled- and triangled-connected lines respectively. **b.** School-size distribution in Abruzzo (○-green) and Tuscany (△-magenta). Both pdf are approximately lognormal and bimodal with splitting point equal to 128 and 151 students per school respectively.

in Abruzzo and only 7 in Tuscany. Only 30% of them have one school in Abruzzo while 80% of them have at least one school in Tuscany. Overall, there are 53 comuni in Abruzzo without schools; only 3 in Tuscany.

Such a differences reflects on the school-size distribution, depicted in Fig. 4.10b. Although primary schools distribute in both regions in terms of size with two peaks, both Abruzzo m_1 and m_2 are shifted on the left w.r.t. the Tuscany ones. The average school-size is smaller in Abruzzo ($\hat{\mu}_{ABR} = 4.56$ ($\hat{\mu}_{ABR}/\ln(10) = 1.98$) versus $\hat{\mu}_{TOS} = 4.91$ ($\hat{\mu}_{TOS}/\ln(10) = 2.13$)), and, remarkably, the lower tail is fatter in the former region. The cutoff for splitting the mixed distributions amounts to 128 in Abruzzo and 151 in Tuscany, and 31% of the schools are clustered in the second peak in the former region; $P(x_i > \bar{\mu}_{TOS} | \forall\, i \in TOS) = 0.38$ in the latter.

In Fig. 4.11a we show, following the same clustering technique used in Fig. 4.8, that in both regions the fraction of big schools within comune k, $P_k(x_i > \bar{\mu} | \forall\, i \in k)$, increases monotonically with respect to the number of inhabitants for $\langle p_k \rangle_h < 20000$.

In this interval, a comparison with figures for entire Italy, plotted in

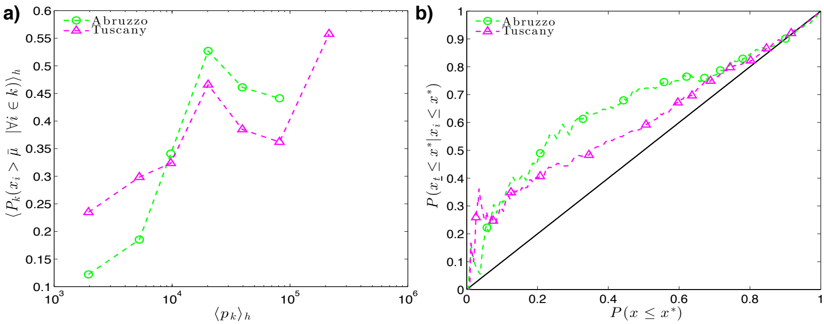


Figure 4.11: Regional analysis II. a. Average fraction of big schools in each comuni bin, defined by Eq. 4.3, in Abruzzo (○-green) and Tuscany △-magenta). The plot shows that more populated comuni are, on average, more likely to have schools sized around m_2, in both regions. Yet, in mountain regions, such as Abruzzo, smaller comuni have also smaller schools on average. **b.** The conditional probability is plotted in the y-axis, for an arbitrary school size x^*, as function of x^* against the cumulative probability $P(x_i \leq x^*)$. The conditional probability is equal to the cumulative in correspondence of the black line. Along these points, there is no attraction between schools of the same size. This is not the case in both the two regions.

Fig. 4.8, reveals that both regions follow the same national pattern. Yet, mountain regions, such as Abruzzo, have a significantly smaller concentration of big schools. In particular in Abruzzo, only about $1/10$ of comuni with just one school, with an average population of roughly 2000, have a school with more than 125 students. In Tuscany, they are the 25%, about the same as national ratio. In larger comuni, with an average population of 5000 and two schools provided (the second bin), the probability of having big schools raises to 0.2 in Abruzzo, still smaller than Tuscany where $\langle P_k(x_i > \bar{\mu} | \forall\, i \in k) \rangle_{h=2} = 0.3$.

Small schools are mainly located in the countryside, and for that reason they cluster together, i.e. it is more likely to find a small school near a small one. In Abruzzo this clustering effect is stronger than in Tuscany. We investigate this point in Fig. 4.11b, where we compute, and plot on the x-axis, the cumulative probability $P(x_i \leq x^*)$, as function of x^*, and the correspondent conditional probability $P(x_{\underline{t}} \leq x^* | x_i \leq x^*)$, on the y-axis, which is the fraction of schools with the size smaller than x^*

among the schools closest to a school of size x^*. This quantity is equal to 74% and 65% for $x^* \equiv \bar{\mu}_{reg}$ in Abruzzo and Tuscany respectively, meaning that there is a greater probability that a small school matches with another of the same kind in the former region. If the conditional probability were equal to the cumulative, as indicated by the black line in Fig. 4.11b, the sizes of neighboring schools would be independent. This is not the case in either the two regions. The probability that a small school has a smaller nearest neighbor is larger than the probability that any school is smaller than a given one. Indeed, the two curves (green for Abruzzo and magenta for Tuscany) are significantly above the 45 degree line for $P(x_i < x^*) < 0.6$ in Tuscany and for $P(x_i < x^*) < 0.7$ in Abruzzo. These probability values roughly correspond to the probabilities $P(x_i < \bar{\mu})$ in respectively Tuscany and Abruzzo, indicating that in both regions small schools are likely to belong to the small mountainous comuni, whose nearest neighbors are of the same class.

We further study the attraction intensity among small schools by disentangling the effect between the countryside and dense zones. To this end, we analyze the GPS location of the schools in the two regions and, for each school i, we compute the number of schools n^i_m belonging within a circle of radius r_m centered at each school j. We exclude from n^i_m all the schools which do not belong to Tuscany or Abruzzo, respectively. To eliminate the effect of region's boundaries, we also compute areas D^j_m as the areas of the intersections of these circles with a given region (Abruzzo or Tuscany). Thus $D^i_m \leq \pi(r^i_m)^2$, because these areas do not include the seaside and administrative territories of other regions. The difference between two subsequent circles yields the area of the annulus $A^i_m = D^i_m - D^i_{m-1}$. The density of schools in the area A^i_m is then defined as:

$$\rho^i_m = \frac{n^i_m - n^i_{m-1}}{A^i_m}, \tag{4.6}$$

and the average density of schools as function of a distance to a randomly selected school is

$$\langle \rho_m \rangle_i = \frac{\sum_N n^i_i - \sum_N n^i_{m-1}}{\sum_N A^i_m}. \tag{4.7}$$

In Fig. 4.12a red lines represent the average school-density around *all* the schools in Tuscany and Abruzzo, which are 472 in the former and 1037 in the latter region. Green lines describe the average school density around a small school with $x_i \leq \bar{\mu}$, named S_1, whereas the blue lines describe the density around large schools, S_2. 64% of the schools in Abruzzo belong to the S_1 group, 53% in Tuscany. Fig. 4.12a collects evidence about the fact that small schools S_1 are located in low school density zones and, accordingly, have a smaller probability to be surrounded by competitor schools than large schools (S_2) located in densely populated areas. In both regions, in fact, the green line goes under the blue one, for at least first $50km$. In particular, within this distance, in Abruzzo the density stays almost constant at approximately 0.053 meaning that 1 school is provided every $20km^2$. In Tuscany, this figure goes up to 0.07, because of a generally higher population density, but yet small.

The correlation coefficients between the school size and the distance to it nearest neighbor are negative in both regions, but the magnitude is quite different, equal to 0.34 in Abruzzo, that is 1.7 times greater than in Tuscany (0.20). To reduce the noise, we proceed by clusterizing schools according to their size. Fig. 4.12b confirms this pattern by showing that small schools have on average more distant nearest schools. We look at the size of each school in both regions, and we define the geodesic euclidean distance between the school i and its nearest neighboring school (which we denote by subscript $\underline{t}$) as $d(x_i, x_{\underline{t}})$. The binning algorithm used is to base 2:

$$l = \{\forall\, i \in [1, \ldots, N] :\, 2^{l-1} \leq x_i < 2^l\}. \tag{4.8}$$

This clusterization yields 8 bins, with different average sizes plotted on the x-axis of Fig. 4.12b. On the y-axis, we plot the average distance between the school i, that belongs to the bin l, and its nearest, i.e. $\langle d(x_i, x_{\underline{t}})\rangle_l$. Each school-bin l is depicted by green circles for Abruzzo and magenta triangles for Tuscany. The average distance between the closest schools decreases with respect to the average school size $\langle x_i\rangle_l$ for $\langle x_i\rangle_l > 32$ in both regions meaning that, in general, small schools are sparser than large schools that are more likely to be located in very dense zones, like cities. The non-monotonic behavior of this quantity for $\langle x_i\rangle_l < 32$ in Tuscany can be explained by the fact that such small schools in Tuscany

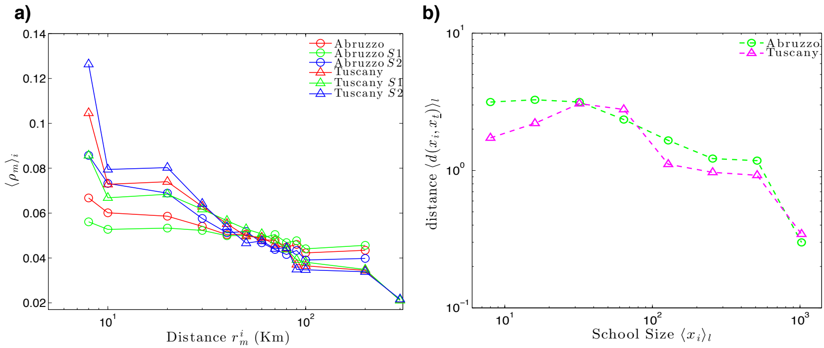


Figure 4.12: Regional spatial analysis. a. $\langle \rho_m \rangle_i$ has been plotted, based on Eq. 4.6, and 4.7, for the region of Abruzzo (□) and Tuscany (△). The red line draws the trajectory averaging among *all* the schools in Italy. Green and blue lines stand for small schools, i.e. $x_i \leq \bar{\mu}$, called S_1, and big schools, i.e. $x_i > \bar{\mu}$, called S_2, respectively. **b.** The average distance, in km, between the closest schools, $\langle d(x_i, x_{\underline{t}}) \rangle_l$, is plotted in Abruzzo (○-green) and Tuscany (△-magenta) with respect to the average size, $\langle x_i \rangle_l$. Each cluster l has been obtained by aggregating schools with near size according to Eq. 4.8. In Tuscany, the schools provided in small islands, at least $20km$ far from the coast, have been removed in order to eliminate any artificial bias from the spatial analysis, whereas the 18% of the schools, with no address provided in the MIUR dataset, have been geocoded in Tuscany according to the GPS localization of the city hall of the comune in which they stand. The average distance between the closest schools decreases in both regions with respect to the average size meaning that, in general, small schools are sparser than large schools that are more likely to be located in very dense zones, like cities.

are usually hospital schools which are located in densely populated areas. Whereas the schools provided in small islands, at least $20km$ far from the coast, have been removed in order to eliminate any artificial bias from the spatial analysis. The three first magenta bins are all below the green ones, confirming, in accordance with the geographical features of the two regions, that in Abruzzo small schools are sparser and more likely to be located in the countryside where the school density is low (see Fig. 4.12a). Moreover, small schools on average have a distance to the nearest neighbor of $4 - 5km$ which is the average distance between a small comune and a more school-dense one (see the Methods section).

The two regions then outline very different patterns of the school system in the countryside. In Abruzzo small schools are uniformly distributed across small comuni, as a result of a policy favoring the disaggregation of the comuni and school organization, due to a tight geographical constraint. In Tuscany, instead, a different system has been implemented, according to geographic features and a higher population density, where small comuni are larger and do not necessarily have small schools, especially if they stand in very populated zones.

4.3 Discussion

We have studied the main features of the size distribution of the Italian primary schools, including the sources of the bimodality, and we have investigated its relation to the characteristics of the Italian cities. The fat left tail of the distribution is the consequences of political decisions to provide small schools in small (mostly countryside) comuni, instead of increasing the efficiency of public transportations. This is most probably caused by the topographical features of the hilly terrain making transportation of students dangerous and costly. The evidence of this conclusions is that hilly cities like Palermo, Napoli, an, above all, Genoa, with steep mountains that end up into the sea, have higher fraction of small schools than mainly flat cities like Torino and Milano.

The analysis of schools growth rates highlights that the schools dynamics follows the Gibrat law, and both the growth rate distribution and the size distribution are consistent with a Bose-Einstein process. Alternatively, the exponential decay of the upper tail can be explained by a constraint by the size of the building or a traveling distance and transportation cost.

Despite our results are conducted using data on Italian primary schools, they predict that schooling organization would be different in another country with different geographical features. Flat territory would lead to open schools in the main villages allowing the children residing in the smallest ones to travel daily. This result is additionally supported by the fact that no territorial constraint has been imposed to the schooling choice. Despite parents can enroll children in the most preferred school, primary students generally do not move *across* comuni to attend a school. Accordingly, we find that school density and school-size are prevalently driven by the population density and then by the geographical features of the territory, as a result of a random process in the school choice made by the parents. This goes in the opposite direction with what has been found in other countries such as USA where school choices influence residential preferences of parents and drive the real estate prices in townships depending on the quality of their schools [30].

The availability of new longitudinal school data will be relevant to a more in-depth analysis and further discussions. Moreover, the availability of data for other similar countries would favor comparison and would be useful to assert our theory. We believe that this study, and future research, can lead to a higher level of understanding of these phenomena and can be useful for a more effective policy making.

4.4 Methods

In this section we propose a novel algorithm for the analysis of spatial distribution of primary schools in entire Italy. This algorithm is needed if the exact coordinates of individual schools are not available, but instead, the centers and the territories of all the comuni are known. For each comune k, we define a gravity center g_k of its territory corresponding to the GPS location of its city hall, and t_k as the area of the comune administration. In Italy the city hall is located in the center of the densely populated part of the administrative division, in order to be easily reachable by the majority of inhabitants. We develop a novel spatial-geographical approach consisting of a sequence of geographic regions bounded by two concentric circles, that we exemplified in Fig. 4.13a for a comune in Abruzzo. First we define a set Z_m^k of comuni whose city halls are within a circle of radius r_m^k and the center at the city hall of comune k. Formally,

$$Z_m^k = \{\forall\, j \in [1, .., K] : d(g_k, g_i) \leq r_m^k\}. \tag{4.9}$$

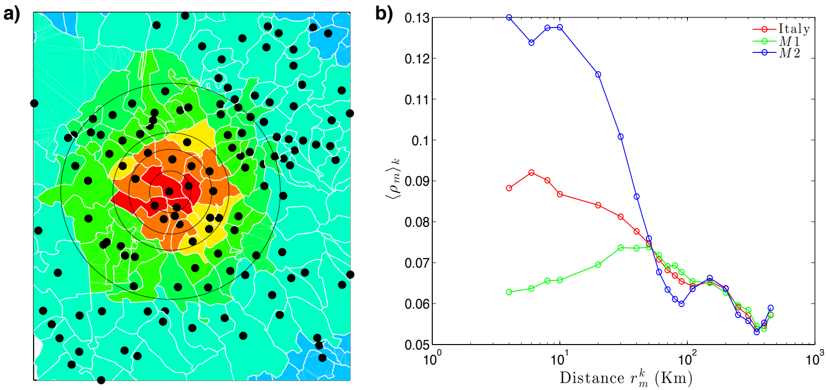


Figure 4.13: Spatial analysis. a. Graphical example for a small comune in Abruzzo of the algorithm used in Fig. 4.13b, based on the Eq. 4.11, 4.12, and 4.13. Different comuni are colored according to the annulus in which they belong. **b.** $\langle \rho_m \rangle_k$ has been plotted for a radius r_m^k of length 10^3 across Italy. The red line draws the trajectory averaging among *all* the cities in Italy. Green and blue lines stand for cities with probability $P_k(x_i > \bar{\mu} | \forall i \in k) \leq 1/2$, labeled $M1$, and $P_k(x_i > \bar{\mu} | \forall i \in k) > 1/2$, labeled $M2$, respectively. Maps generated with Matlab.

Next we compute the number of schools provided by the comuni which are members of set Z_m^k that is defined by

$$n_m^k = \sum_{j \in Z_m^k} n_j \tag{4.10}$$

and their area

$$D_m^k = \sum_{j \in Z_m^k} t_j, \tag{4.11}$$

where t_j is the area of comuni j. Next we compute the area associated with all the comuni in the m-th concentric annulus surrounding comune k as the difference between the area associated with the larger circle m of radius r_m^k and the area associated with the smaller circle $m-1$ of radius r_{m-1}^k, i.e. $A_m^k = D_m^k - D_{m-1}^k$. In Fig. 4.13a, each comune territory is colored with different colors according to the annulus in which they belong.

The density of schools in the area A_m^k is then defined as:

$$\rho_m^k = \frac{n_m^k - n_{m-1}^k}{A_m^k} \tag{4.12}$$

Then we compute the average density of schools around any school in Italy as:

$$\langle \rho_m \rangle_k = \frac{\sum_K n_m^k - \sum_K n_{m-1}^k}{\sum_K A_m^k} \tag{4.13}$$

In Fig. 4.13b, we plot $\langle \rho_m \rangle_k$ averaged over all the $K = 8092$ Italian comuni as a function of the radius r_m that goes up to 10^3 Km across the entire Italy. The red line represents the average school-density among *all* the cities in Italy. On average, Italian comuni stand within very dense zones providing almost 1 school per $10km^2$. The dense zones generally last for $10km$ and, after that, a smoothed depletion zone is experienced. However, the average distance between a comune k and a very large city with many schools is about $100km$, accordingly we see a second peak in the average school density at distance $100km$.

The full sample analysis basically averages heterogeneous characteristics that feature different types of comuni. The interaction among schools can be better understood by splitting the sample according to $P_k(x_i > \bar{\mu}|\forall i \in k)$. In Fig. 4.13b, comuni with $P_k(x_i > \bar{\mu}|\forall i \in k) \leq 1/2$, i.e. with predominantly small schools, are named $M1$. The others, with predominantly big schools, are called $M2$.

- $M2$-comuni, the blue line, are (on average) more likely to be surrounded by school-dense cities. They are cities located in densely populated areas (depicted in red in Fig. 4.9) where the school density is large (1.3 schools stand on average within $10km^2$). As far as the distance increases mountainous areas (and hence $M1$-comuni) are encountered and, as a result, the density of schools is found to dramatically decrease.

- The green line describes instead cities labeled $M1$ where a smaller school density is found. Within $10km$, in fact, almost 1 school every $20km^2$ are encountered on average, about the half of what we find for the $M2$-comuni. This is because $M1$-comuni mainly stand along the countryside (those depicted in blue in Fig. 4.9) where school density slowly increases with distance and reach a maximum at approximately $40km$, which can be interpreted as a typical distance to a densely populated area in a neighboring mountain valley. After this distance the density of schools around $M1$ and $M2$ comuni behave approximately in the same way.

4.5 Supplementary Information

4.5.1 Italian private primary schools versus public primary schools: a comparison.

In this chapter we addressed the source of the bimodality by considering all the Italian primary schools. Here we focus on the potential effect of school type on the school-size distribution. Our dataset collects $N = 17,187$ primary schools in Italy. The fraction of private schools was always low during the past century. In Italy only the 9% of the total of primary school are private.

The main source of primary school privatization within the country is religion. Most of the private schools are venues where education is strictly connected with the Catholic confession. Among the private schools more than 73% are of Catholic inspiration. Straightforward historical roots are expected to explain the location of the Italian Catholic private schools and only marginal are the geographical reasons: private schools are in fact only the 6.54% of the mountain schools.

We define $\mathcal{M}$ the set of comuni k that are in mountains that, according to the Law n. 991/1952, are those that have at least the 80% of their territories above the 600 meters above the sea and an altitude gap between the higher and the lower point not least than 600 meters. Each comune k has n_k schools and a fraction of private schools in this comune defined as $P(i \in \mathcal{P} | \forall i \in k) \equiv \eta_k$, where i is the school ID. We also define the school-size of a private school i that resides in a mountain comune as $x_{i \in \mathcal{P}, \mathcal{M}}$. Analogously, $x_{i \in \bar{\mathcal{P}}, \bar{\mathcal{M}}}$ stands for the size of a public school residing in a

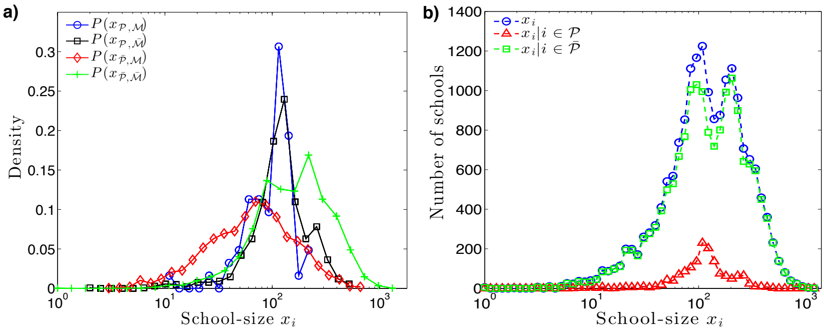


Figure 4.14: **a.** Italian primary school-size distribution disentangled by school type (private, $\mathcal{P}$, versus public, $\bar{\mathcal{P}}$) and geography (mountain, $\mathcal{M}$, versus non-mountain, $\bar{\mathcal{M}}$). **b.** Italian primary school-size distribution by school-type. The blue pattern replicates Fig. 4.1a.

non-mountain comune.

Figure 4.14a shows that neither private mountain schools $(\mathcal{P}, \mathcal{M})$ nor private schools that reside in flat territories $(\mathcal{P}, \bar{\mathcal{M}})$ seem to contribute significantly to the left tail of the school-size distribution. Both the $(\circ)$ blue and the $(\square)$ black lines, respectively, depict two relatively narrow school-size distributions around 100 students per school, the $(+)$ green $(\bar{\mathcal{P}}, \bar{\mathcal{M}})$ and the $(\diamond)$ red lines $(\bar{\mathcal{P}}, \mathcal{M})$. In accordance with the results shown in the main text, mountain public schools mostly contribute to the left tail of the distribution. Finally, the distributions of private schools both for mountain and flat regions are almost identical even though there are only 449 mountain private schools and one might expect large statistical uncertainty.

Figure 4.14b draws the school-size distribution without considering geography but only distinguishing with respect to the school-type. Frequencies are then shown for private (red $\triangle$) and public (green $\square$) schools and compared with the distribution of all the Italian primary schools (in blu $\circ$) that replicates Figure 1a in the main text. It confirms that private schools play only a slight role in generating the left peak which is still present in the the size distribution of public schools.

Figure 4.15a plots the fraction $\eta_c = n_{\mathcal{P},c}/n_c$ of private schools among

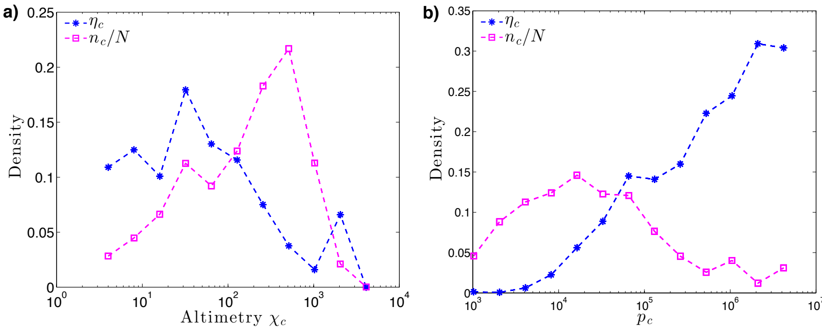


Figure 4.15: Private schools analysis II. a. The fraction of private schools, $\eta_c = n_{\mathcal{P},c}/n_c$ among all schools in bin c with a given altitude above the sea level, χ_c, with respect to the total number of schools n_c in that bin ($*$, blue lines). The distribution of all the schools among the altitude bins n_c/N ($\square$, magenta lines). **b**. The fraction of private schools $\eta_c = n_{\mathcal{P},c}/n_c$ among all schools in bin c, formed by comuni with a given number of inhabitants, p_c. As a robust check we also plot in ($\square$) magenta the distribution of schools (both private and public) in each bin c, n_c/N, versus the population.

all schools (public and private) in each bin c of comuni with given altitude, against their altitude above the sea level, χ_c. In order to reduce the noise, we binned comuni according to the altimetry:

$$c = \{\forall\, k \in [1, \dots, K] : 2^{c-1} < \chi_k \leq 2^c\}. \tag{4.14}$$

It yields 11 bins, $c \in [1, \dots, 11]$, each of them collecting comuni according to the meters above the sea level. The figure also shows the distribution n_c/N of all the schools among the elevation bins. The figure provides evidence that the majority of private schools (in $*$ blue) are located in the comuni with low altitude $\chi_c < 128m$. In contrast the distribution of number schools with given altitude (both private and public, in $\square$ magenta) reaches the maximum for comuni with altitude $\chi_c = 512m$. The hill-shape of this distribution can be explained by the unequal territory covered by different bins. We conclude that there is a greater fraction of private schools in the planes than in the mountains.

Finally, using the same binning algorithm in Eq. 4.4, Fig. 4.15b shows strong positive correlation between the fraction of private schools $\eta_c =$

$n_{\mathcal{P},c}/n_c$ among all schools in bin c, η_c, of comuni with given number of inhabitants, p_c (in $*$ blue), confirming that the location of the Italian Catholic private schools mainly roots in the more populated comuni.f As a robust check we also plot in ($\square$) magenta the distribution, n_c/N, of all schools (both private and public) among population bins c, which, consistently with the analysis of Fig. 4.4-4.5, yields the Italian city-population distribution which has a slightly skewed shape. In very small comuni ($p_c < 10^4$), where a greater quantity of schools is provided, we count only a small fraction of private ones. Conversely, the fraction of private schools in the large comuni is very large (e.g. private schools constitute 30% of all schools located in Rome, in comparison to the 9% nationwide).

Large flat comuni are then very likely to be the places where most of Italian private primary schools are located. We conclude that privatization has been driven across the years for religious confessional purposes rather than following the unmatched education demand in the countryside due to the lack of the public system.

4.5.2 Testing unimodality in the school-size distributions of flat comuni.

In this section we address concerns on bimodality on the school-size distribution of flat comuni. In the section 4.2.2 we have demonstrated that geography is the main source of bimodality in the school-size pdf showing that mountain schools clusterize around m_1. Yet there might be other confounding factors that might keep a second peak, i.e. m_1, in the school-size pdf of the schools that reside in flat comuni.

In Fig. 4.6b we distribute schools according to the number of students, x_i, conditional on the altimetry of comuni. As we discuss in the section 4.2.1 this analysis gives five distributions which correspond to different elevation bins. The PDFs of mountain schools stand on the left and on the right we have flat schools. The ($\circ$) green line shows the school-size distribution for $N_{250m} = 3,033$ schools that reside in comuni with around 250 meters from the sea level. Despite this PDF does not show a sharp peak corresponding to m_2, and thus potentially might be bimodal, here we demonstrate that *statistically* the hypothesis in favor of unimodality can not be rejected.

To see that we use the complementary error function to estimate the probability that the number of schools in the central bin n_1 is not significantly smaller than and the numbers of schools in the neighboring two bins n_2, n_3 are not significantly larger than a certain number n^* provided that the standard deviation of the number of schools in these bins due to small statistics is $\sqrt{n^*}$:

$$p(n^*) = \frac{1}{2}\Pi_i \text{erfc}\left(\frac{|n_i - n^*|}{\sqrt{2n^*}}\right) \tag{4.15}$$

This is equivalent to test the hypothesis that the distribution is unimodal. In the school-size distribution for schools that reside in comuni with around 250 meters above the sea, the central bin collects $n_1 = 639$ schools. On either sides there are two other bins that collect $n_2 = 670$ and $n_3 = 646$ respectively. The probability that the distribution is not bimodal is maximum for $n^* = 646$ where it is equal to $p_{max}(n^* = 646) = 0.15$. Fixing a level of confidence of 0.10 we, therefore, cannot reject the hypothesis of unimodality.

Bibliography

[1] Gabaix, X., Power Laws in Economics and Finance., *Annu. Rev. Econ.* **1**, 255–93, (2009). 105

[2] Gabaix, X., Zipf's Law for Cities: An Explanation., *Q J Econ.* **114**, 739–67, (1999). 105, 113, 115

[3] Allen, P.M., *Cities and regions as self-organizing systems: models of complexity*, (Routledge, 1997). 105, 113, 115

[4] Amaral, L. A. N., et al., Power Law Scaling for a System of Interacting Units with Complex Internal Structure., *Phys. Rev. Lett.*. **80**, 1385–1388 (1998). 105

[5] Byrne, D., *Complexity theory and the social sciences: an introduction*, (Routledge, 2002). 105

[6] Caves, R. E., Industrial Organization and New Findings on the Turnover and Mobility of Firms., *J Econ Lit.* **36**, 1947–82, (1998). 105

[7] Bak, P., *How nature works*, (Oxford University Press, 1997). 105

[8] Kauffman, S., *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity: The Search for the Laws of Self-Organization and Complexity.*, (Oxford University Press, 1996). 105

[9] Gazzetta Ufficiale, *Decreto del Presidente della Repubblica del 20 marzo 2009 n. 81*, (2 luglio 2009). 105

[10] Disposizioni concernenti la riorganizzazione della rete scolastica, la formazione delle classi e la determinazione degli organici del personale della scuola., *Decreto Ministeriale* **331**, (24 luglio 1998). 105

[11] Belmonte, A. & Pennisi, A., Education reforms and teachers needs: a longterm territorial analysis., *IJRS.* **12**, 87-114 (2013). 105

[12] De Wit, G., Firm Size Distributions: An Overview of Steady-State Distributions Resulting form Firm Dynamics Models., *Int J Ind Organ.* **23**, 423–50 (2005).

[13] Fu, D. et al., The growth of business firms: Theoretical framework and empirical evidence., *Proc Natl Acad Sci USA.* **102**, 18801–18806 (2005). 106, 110

[14] Gazzetta Ufficiale, *Decreto del Presidente della Repubblica dell'8 marzo 1999 n. 275*, (10 Agosto 1999). 108

[15] Pitman, E. J. G., The estimation of the location and scale parameters of a continuous population of any given form., *Biometrika* **30**, 391–421 (1939). 109

[16] Stanley, M. H. R. et al., Zipf plots and the size distribution of firms., *Econ Lett.* **49**, 453–457 (1995). 109

[17] Gibrat R., *Les Inégalités économiques*, (Recueil Sirey 1931).

[18] Sutton, J., Gibrat's Legacy., *J Econ Lit.* **35**, 40–59 (1997).

[19] Fu, D. et al., A Generalized Preferential Attachment Model for Business Firms Growth Rates-I. Empirical Evidence., *Eur Phys J B.* **57**, 127–130 (2007).

[20] Pammolli, F. et al., A generalized preferential attachment model for business firms growth rates: II. Mathematical treatment., *Eur Phys J B.* **57**, 131–138 (2007). 110, 111

[21] Axtell, R. L., Zipf Distribution of U.S. Firm Sizes., *Science* **293**, 1818–20 (2001). 110

[22] Growiec, J., Pammolli, F., Riccaboni, M. & Stanley, H.E., On the Size Distribution of Business Firms., *Econ Lett.* **98**, 207–12 (2007). 110

[23] Stanley, M. H. R., et al., Scaling behaviour in the growth of companies., *Nature* **379**, 804–6 (1996). 110

[24] Ayebo, A. & Kozubowski, T.J., An asymmetric generalization of Gaussian and Laplace laws., *J. Probab. Stat. Sci.* **1**, 187–210 (2003). 111

[25] Buldyrev, S. V., Pammolli, F., Riccaboni, M., & Stanley, H. E., *The Rise and Fall of Business Firms*, To be Published (2014). 112

[26] Wang, J., Wen, S., Symmans, W. F., Pusztai, L. & Coombes, K. R., The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data., *Cancer Inform* **7**, 199–216 (2009). 113

[27] Clauset, A., Shalizi, C. R. & Newman, M. E. J., Power-law distributions in empirical data., *SIAM review.* **51**, 661–703 (2009). 113

[28] Eeckhout, J., Gibrat's Law for (All) Cities., *Am Econ Rev.* **94**, 1429–51 (2004). 113, 115

[29] Reed, W. J., The Pareto, Zipf and Other Power Laws., *Econ Lett.* **74**, 15–9 (2001). 113, 115

[30] Black, S. E., Do better schools matter? Parental valuation of elementary education., *Q J Econ.* **114**, 577-599 (1999). 128