PhD THESIS DECLARATION

I, the undersigned

FAMILY NAME	Orlando
NAME	Tommaso
Student ID no.	1413699

Thesis title:

Institutions and organizations in economic perspective

PhD in	Economics and Finance
Cycle	28th
Student's Tutor	Prof. Maristella Botticini
Year of thesis defence	2018

DECLARE

under my responsibility:

1) that, according to Italian Republic Presidential Decree no. 445, 28th December 2000, mendacious declarations, falsifying records and the use of false records are punishable under the Italian penal code and related special laws. Should any of the above prove true, all benefits included in this declaration and those of the temporary "embargo" are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree no. 224, 30th April 1999, to keep a copy of the thesis on deposit at the "Bilioteche Nazionali Centrali" (Italian National Libraries) in Rome and Florence, where consultation will be permitted, unless there is a temporary "embargo" protecting the rights of external bodies and the industrial/commercial exploitation of the thesis;

3) that the Bocconi Library will file the thesis in its "Archivio Istituzionale ad Accesso Aperto" (Institutional Registry) which permits online consultation of the complete text (except in cases of temporary "embargo");

4) that, in order to file the thesis at the Bocconi Library, the University requires that the thesis be submitted online by the student in unalterable format to Società NORMADEC (acting on behalf of the University), and that NORMADEC will indicate in each footnote the following information:

- PhD thesis: Institutions and organizations in economic perspective;

- by: Orlando Tommaso;
- defended at Università Commerciale “Luigi Bocconi” – Milano in the year: 2018 ;
- the thesis is protected by the regulations governing copyright (Italian law no. 633, 22nd April 1941 and subsequent modifications). The exception is the right of Università Commerciale “Luigi Bocconi” to reproduce the same, quoting the source, for research and teaching purposes;

5) that the copy of the thesis submitted online to Normadec is identical to the copies handed in/sent to the members of the Thesis Board and to any other paper or digital copy deposited at the University offices, and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (Italian law, no. 633, 22nd April 1941 and subsequent integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal, and shall be exempt from any requests or claims from third parties;

7) that the thesis is not subject to “embargo”, i.e. that it is not the result of work included in the regulations governing industrial property; it was not written as part of a project financed by public or private bodies with restrictions on the diffusion of the results; is not subject to patent or protection registrations.

Date: 26/01/2018

Abstract

This volume is composed of four research papers dealing with the microeconomic analysis of historical and present-day, private and public, formal and informal institutions.

Chapter 1 ('Apprenticeship and learning in modern England') looks back at modern-age apprenticeship and provides an interpretation of this institution as a learning tool. Such interpretation allows us to reconcile bits of historical evidence concerning the evolution of apprenticeship on the eve of the Industrial Revolution.

Chapter 2 ('Resistance to institutions and cultural distance: brigandage in post-unification Italy') analyzes patterns of violent unrest that occurred during the Italian national unification to study how cultural distance affects the effectiveness of institutional transplantations. It shows that propensity to revolt against new institutions is weaker if the receiving environment shares some of the donor's cultural traits.

Chapter 3 ('Corruption and personnel selection and allocation in the public sector') studies the effects of corruption on the composition of the public workforce and on the efficient allocation of human resources within the public sector. Using Italian local-level data, the paper highlights a relative reduction in human capital, as well as a relative increase in under-education, among public employees in areas with higher levels of corruption.

Chapter 4 ('M&A negotiations with limited information: how do opaque firms buy and get bought?') deals with mergers and acquisitions of firms, giving a theoretical underpinning to a body of empirical evidence concerning the relationship between the difficulty to assess firms' values and observable characteristics of M&A interactions, such as the method of payment and the bid premium.

Acknowledgments

My most sincere thankfulness goes to my advisor, M. Botticini, for providing me with continued assistance and support, within and beyond research endeavors. During these years I have also greatly benefited from the advice and cooperation of several members of the faculty at Bocconi University: in particular, P. Battigalli, F. Maccheroni, M. Marinacci, N. Pavoni and G. Tabellini. Much of the work presented here must be credited to the efforts of my co-authors. None of it would exist without the invaluable help and support of my PhD classmates and co-students, of colleagues at the Bank of Italy, and – most especially – the intersection thereof. While not appearing explicitly, the contributions of Veronica and Paolo lie underneath every line. I owe all achievements to the love and care of my family and Valentina.

Contents

Part of the research contained in this volume was carried out while the author worked at the Bank of Italy. The views expressed herein are the author's only and do not necessarily reflect those of the Bank of Italy.

Introduction

This volume is composed of four research papers dealing with the microeconomic analysis of historical and present-day, private and public, formal and informal institutions. Several aspects contribute to make these papers mutually interrelated. Two papers (Chapters 1 and 2) explicitly take on a historical point of view, referring to aspects of modern and contemporary institutional and economic history that had a relevant impact on economic and political transformations. The same papers also share a special focus on the long-run transmission of individual traits. Two papers (Chapters 2 and 3) deal with phenomena concerning public institutions and illegal activities: the dynamics of violence following the transplantation of an administrative and legal apparatus; and the relationship between corruption and the composition of the public workforce. These two papers also share a special focus on the Italian case and use Italian local-level data. The remaining two papers (Chapters 1 and 4) deal with institutionalized market agreements that bear particular relevance for their aggregate effects on – respectively – historical labor markets (Chapter 1) and the present-day market for corporate control (Chapter 4). Finally, all papers except Chapter 4 share a more or less explicit attention towards the important role played by cultural values and attitudes (towards the law, the government and the market) in shaping economic outcomes.

More specifically, Chapter 1 ('Apprenticeship and learning in modern England') contains an economic analysis of historical *apprenticeship*, a labor-market institution that gained prominence in many European countries throughout the Middle Ages and became a mass phenomenon in modern-age England. As recently noticed by De la Croix et al. (2017), apprenticeship was inextricably related to professional *guilds*. Economic historians have proposed different interpretations of the economic role of apprenticeship – besides its obvious training function – based on its salient characteristics of costliness, compulsoriness and extended duration. On the one hand, apprenticeship has been viewed as one of the

barriers to entry raised by guilds;[1] on the other, it may have functioned as second-best agreement to overcome a problem of misaligned incentives between the master and the young worker (Wallis, 2008); finally, apprenticeship could have constituted a means for the worker to signal his/her ability to competing potential employers (Tadelis and Rangel, 2001).

All of these theories provide insightful explanations of central aspects of modern-age apprenticeship, but none of them appears to be entirely exhaustive of the economic role of this institution, in that they seem to collide with pieces of historical evidence available to us. The barrier-to-entry theory is hardly reconcilable with the fact that many apprentices were sons and daughters of guild masters, unless one believes that guilds had ended up in a very unpleasant equilibrium despite their ability to coordinate. The hold-up problem, whose solution would require an extended period of forced supply of cheap labor by the apprentice, was unlikely to arise in less technically sophisticated professions, where the little amount of training actually provided to apprentices did not require any particular investment by the master. Finally, exactly because most apprentices were left for many years with little role within the master's shop, apprenticeship looks like a deeply inefficient way for workers to signal their ability.

In this paper, an interpretation of apprenticeship as a learning tool is advanced, theoretically motivated and supported by some suggestive historical evidence. Specifically, a model in which apprenticeship allows agents to gain information about unobservable individual traits is developed. These traits (entrepreneurial attitudes) are important determinants of individual success in risky activities and cannot be inferred but through the practice of the activity itself. Beliefs about these traits are subject to intergenerational transmission: altruistic parents would send their kids to become apprentices in order to let them learn not only about a specific craft, but about themselves. What is learned during the apprenticeship period then becomes the subsequent generation's initial belief. When demand for apprenticeship is sufficiently high, current entrepreneurs have an incentive to provide this form of training in exchange for a price and to sustain an institutionalized agreement (the guild) limiting the number of intakes. As dynasties progressively accumulate information about their own traits, the demand for this kind of training decreases, as does the average price paid by the apprentices. When profits from apprenticing become too small, the masters have no incentives to sustain the guild, whose dismissal removes the intake limitations but produces a sudden increase in prices faced by apprentices.

[1] This opinion has been advanced within the debates among economic historians around guilds in the Modern Age: see, for instance, Ogilvie (2004) and Epstein (2008) for accounts of the two competing positions.

The predictions of this simple model are in line with three distinct pieces of historical evidence. First, the model explains why apprenticeship was chosen both by 'outsiders' (i.e. young boys and girls migrating from the countryside or whose parents worked in a non-guilded profession) and by the children of guild masters. Second, the model satisfactorily tracks the evolution of apprenticeship premiums in 17th and 18th century, accounting for its gradual decrease and subsequent rapid rise to a more or less constant level. Finally, the model makes the gradual decrease in the number of apprenticed workers, culminating in the final demise of historical apprenticeship, a built-in characteristic of this institution.

Chapter 2 ('Resistance to institutions and cultural distance: brigandage in post-unification Italy', co-authored with G. Lecce and L. Ogliari) is a study on the relationships between cultural traits and institutions. More specifically, we analyze whether the short-run outcomes of *institutional transplantations* can be affected by variations in cultural traits within the so-called *receiving environment*, i.e. the socio-economic context upon which the institutions are imposed. In particular, we investigate whether different levels of *cultural distance* between recipient communities and the *donor environment* (i.e. the polity that implements the transplantation) are related to different degrees of acceptance of the transplanted institutions. *Institutional acceptance* represents the first necessary condition for a successful institutional transplantation: if the transplant is rejected, the new institutions will never be at work in the recipient environment.[2]

In this paper we exploit an instance of institutional transplantation that occurred in the context of the Italian unification (1860-1870), to focus on the relationship between cultural distance and institutional acceptance. Starting in 1860, continental Southern Italy was incorporated into the nascent Italian kingdom and received the transplantation of the whole administrative and legal apparatus of Piedmont, the polity which led the unification process and the donor environment in this context. While a similar transplantation occurred in the rest of Italy as well, continental Southern Italy presents two peculiar features which make it an ideal setting for our study. First, in the South the reaction against the new institutions took the form of a violent guerrilla-like movement known as *brigandage*, whose intensity we use to measure the degree of institutional rejection. Second, continental Southern Italy is characterized by the presence of communities settled in the Middle Ages by migrants originating from Piedmont and nearby areas. These 'near-Piedmontese' communities inherited the cultural traits of their ancestors and were thus culturally closer

[2]Lecce and Ogliari (2017) address a related but distinct question: using the Napoleonic occupation of German counties, they show that cultural distance hinders *institutional sedimentation*, i.e. the possibility of transplanted institution to function in an effective way, once established.

to Piedmont than any other Southern community. Besides having persisted within near-Piedmontese settlements, we hypothesize that Piedmontese cultural traits have undergone a geographically limited but socially non-negligible process of diffusion, spreading to neighboring communities as well. In this case, we can use geographic distance from each municipality in continental Southern Italy to the closest near-Piedmontese settlement as a proxy for the cultural distance between that municipality and Piedmont.

By means of secondary data digitized by us, we construct a novel database listing, for each municipality in continental Southern Italy, all recorded episodes of brigandage that took place there between 1860 and 1870. Controlling for a number of geographic, demographic and economic variables, we find a strong positive association at the municipal level between the intensity of brigandage and the distance from the nearest near-Piedmontese community. This results is robust to a wide array of specifications, including the use of a network of ancient Roman roads to compute approximate, but historically more accurate, road-distances between Italian municipalities in 1860-1870. As a confirmation of our diffusion theory, we find the result to be driven by variations within a 50-kilometer radius around near-Piedmontese municipalities, that is the distance within which frequent contacts are likely to have taken place. Furthermore, we address the possibility that results are driven by proximity to linguistic enclaves, irrespective of their being of Piedmontese descent. We are helped by the presence of several other enclaves in continental Southern Italy: repeating our exercise using these as the reference points, we find no significant effect. Finally, we also show that brigandage had a negative effect on electoral turnout in the early electoral rounds in Italian history, using a difference-in-differences approach to control for other potential determinants of political participation.[3]

Chapter 3 ('Corruption and personnel selection and allocation in the public sector', co-authored with S. Mocetti) presents evidence on the relationship between corruption and the quality of the public workforce. Several authors have so far focused both on the impact of corruption on the functioning of the public sector (e.g. Hessami, 2014) and on the association between individual dishonesty and the likelihood of subjects to self-select into public employment (e.g. Hanna and Wang, 2017). We contribute to this literature by investigating whether areas characterized by higher corruption levels show systematic differences in (i) the composition of the public-sector workforce in terms of human capital and (ii) the quality of human resource management within public administrations.

To this end, we combine two rich sources of information about Italian skilled workers and corruption intensity at the local level. Data about individual workers are obtained

[3]For a recent paper relating Italian brigandage and political participation, see Accetturo et al., 2017).

from the Labour Force Surveys administered by the Italian National Statistical Institute (Istat). Data on corruption and related crimes are provided by the Italian Ministry of the Interior. We aggregate crime data at the Local Labor Market (LLM) level, a statistically motivated geographic partition of the Italian territory, which is meant to maximize the likelihood that any individual living in a given LLM also works in the same LLM. We then use a difference-in-differences approach using the level of corruption as treatment and public employees and workers of the manufacturing sector as, respectively, the treatment and control groups.

On average, controlling for a number of individual and LLM-specific characteristics, Italian public employees turn out to be more educated than their private counterparts. We first document that higher levels of corruption are associated with a reduction in this human capital advantage, which is particularly evident for workers in top positions (managers and highly qualified professionals). This may be due either to highly educated workers choosing not to apply for public sector jobs where these are more likely to be affected by corruption (i.e. to a distortion in *self-selection*) or to public agencies in corrupt areas losing their ability to recruit workers according to objective criteria related to human capital endowments (i.e. to a distortion in *screening*).

We also find that the allocation of public employees across tasks is worse, in relative terms, where corruption is more widespread. In particular, we compare shares of *under-educated* workers, identified in our data as those whose educational attainments lie below the 25th percentile of the distribution of attainments in their professional category. Corruption is related to an increase in the rate of under-educated workers in the public sector with respect to the manufacturing sector. In aggregate terms, descriptive evidence suggests that the under-education net advantage the public sector shows in low-corruption LLMs is reversed into a net disadvantage in high-corruption LLMs. Once again, the effect is much larger for managers and high-ranking professionals. Moreover, we show that hours worked and absence rates also suffer from the detrimental impact of corruption.

Our results are robust to several alternative specifications, aimed at addressing the most obvious identification issues. First, corruption may be the consequence of relatively worse selection and allocation patterns within the public sector. To deal with this difficulty, we provide IV estimates of our parameters of interest, instrumenting current corruption by spells of historical domination by regional and external polities (from Di Liberto and Sideri, 2015) and a measure of dependence of the local productive structure from the public sector, dating back to the 1970s. Second, we show that our results are not driven by specificities of southern LLMs, by repeating our exercise using central and northern LLMs only. We support our claims by showing that the inclusion of possible confounding factors

(such as local unemployment, demography and firm productivity) does not substantially alter our results.

This evidence implies a suggestion for future research in this area as well as a relevant policy message. Indeed, on the one hand we highlight a new channel through which corruption may reproduce itself and become persistent: corruption lowers the quality of the public workforce – and, presumably, the functioning of the public sector – which, in turn, favors incentives towards further corrupt behaviors. In a policy perspective, our work casts some doubts on the efficacy of policies aimed at letting single public bodies check on their exposure to corruption by themselves: if corruption worsens the quality of the public workforce, it may also reduce the ability of public-sector managers (irrespective of their personal honesty) to counteract corruption within their own agencies.

Chapter 4 ('M&A negotiations with limited information: how do opaque firms buy and get bought?', co-authored with P. Battigalli, C. Chiarella and S. Gatti) deals with a long-lasting, fundamental question in the area of M&A studies: how do informational limitations affect the observable outcomes of M&A negotiations? More specifically, we investigate how *firm opacity* – a measure of how difficult it is for a counterpart to give a precise assessment of a firm's value – is related to the two most visible characteristics of real-words acquisition bids: the method of payment (cash or stock) and the bid premium (the amount paid by the bidder in excess of the target's declared market value). These issues have often been addressed theoretically (e.g., among others, in the seminal work by Hansen, 1987) and empirically (see, for instance, Chemmanur et al., 2009). Our work aims at providing this body of literature with a simple yet insightful and sufficiently comprehensive theoretical analysis, yielding testable predictions. After developing our model, we take its predictions to the data, using M&A transactions having occurred between 1985 and 2014 among U.S. publicly listed firms.

We model the M&A interaction as a two-person, two-sided private information signaling game. Our analysis improves over that of Hansen (1987) in one crucial dimension: we allow the *target* firm to have a non-degenerate posterior over the type of *bidder* it is facing, upon receiving a bid. This allows us to expand the range of equilibria that arise within our game. In particular, stock offers can be observed in the context of a separating equilibrium à la Hansen (1987) as well as in the context of a pooling equilibrium in which all types of bidder advance the same stock offer. Our model also allows for multiple equilibria at a single parameter configuration.

We define a player's opacity as the uncertainty implied by the other player's belief about the former's type. Through numerical simulations, we show that our model provides sharp

prediction on the correlation between observables and the opacity of the target: the latter is related to a larger *ex ante* probability of observing a stock offer as well as to larger bid premiums. The intuition behind these correlations is clear, in that target opacity makes offering stock relatively more convenient when all type of bidder converge on the same stock offer: the consequent stock-pooling equilibrium is – by definition – made of stock offers only and also yields the highest bid premiums. The predictions concerning the opacity of the bidder are harder to grasp by intuition, but show a (possibly weak) negative correlation between bidder opacity and the use of stock offers and no correlation (or, possibly, a weakly negative one) between bidder opacity and bid premiums.

For a preliminary empirical evaluation, we make use of a measure of opacity by Bharath et al. (2009), stemming from a principal component analysis of several micro-structural measures of adverse selection risk. We then use a simultaneous equations model inherited from Boone and Mulherin (2007) to estimate the effects of firm opacity on the joint determination of the method of payment and the bid premium in a sample of around 1,650 M&A transactions. Our model's prediction on target opacity are confirmed, although it is important to notice that the relationship between target opacity and the use of stock only appears for transaction with a sufficiently large deal materiality. As for bidder opacity, we find a statistically weak negative association with the use of stock and an unexpected more stable negative association with bid premiums. The discrepancies with some of our model's predictions are likely to arise from the assumption of symmetrically and identically distributed prior beliefs, which does not allow us to take into account empirical regularities such as, for instance, the inverse relationship between size and opacity observed in the data.

Chapter 1

Apprenticeship and learning in modern England

1.1 Introduction

1.1.1 Why apprenticeship?

Can professional education teach something beyond mere work-related operational skills? In particular, can professional training help young individuals discover something about *unobservable* characteristics of theirs whose knowledge would enable them to make wiser choices in their adult future? In order to answer these questions, we focus our attention on the educational institution of *apprenticeship* as it existed in Western Europe during the Modern Age (16th to early 19th centuries). The main motive behind this choice is that the institution of apprenticeship made professional education a mass phenomenon in modern-age Europe, unlike it has ever been in previous and subsequent ages. Moreover, in most European countries and especially in England professional training was subject to explicit regulation by local and national laws. Its degree of formalization can be compared to that of today's schooling and academia.

At that time, schooling was still a prerogative of the élites and, even for them, scholarly education was often imparted by privately hired mentors and clergy, which leaves us with few evidence but narrative accounts. Universities had long been established, but despite being the engines of high culture, their role had not changed much since the Middle Ages: academic curricula aimed at reproducing an intellectual élite of philosophers, theologians and jurists. High-end medics were being trained in universities, but everyday medical and surgical practices were still performed by barbers and dentists who had never attended

college. Most urban tasks that were later to become prerogative of law graduates were performed by individuals who had learned the practice while serving as apprentices and whose scholarly education was often limited to being literate. Thanks to the ongoing revolution in scientific methods, scientific education would gradually become an essential part of the philosopher's curriculum, but the distance between academic science and industrial applications is testified by the fact that many of the radical innovators and inventors who were active at the height of the first Industrial Revolution (in the second half of the 18th century) did not own any academic title.

The fact that professional education was regulated by the law allows us to investigate it in depth, thanks to a fragmentary yet wide array of sources ranging from court records to tax registers. Only recently have similar forms of professional education, very often still under the "apprenticeship" label, received a similar degree of recognition by regulators. Even in countries where professional education is currently widespread and has great social importance, such as Germany, its quantitative relevance seems marginal if compared to its diffusion in the Modern Age.

It is to be believed that modern-age apprenticeship, as an instance of a formalized professional training institution, can be fruitfully used to understand the nature of such institutions' part in the accumulation of knowledge within and throughout generations. This paper builds around the notion that, besides accomplishing the necessary transmission of operational skills, professional education can help agents gain information about unobservable individual traits which constitute important determinants of individual success in risky activities but which cannot be inferred but through the practice of the activity itself. We model the educational and occupational choices of parents that have to decide (1) whether to send their children to be trained as apprentices and (2) whether to engage in a risky, "entrepreneurial" activity or whether to chose a riskless, "traditional" occupation. By transmitting individual characteristics and opinions about those to their kids, parents reproduce the aggregate structure of beliefs in the population. At the same time, education allows young agents to change their opinions, by inferring information about their own traits from meaningful signals. Altruistic parents understand the profitability of education for their children and compare it to the costless preservation of opinions. In what follows we show that a model of this kind can generate patterns of education, in the form of apprenticeship, reflecting some of the actual characteristics of such institution. For instance the model predicts that parents from both occupations send their children to be apprenticed. This is something that appears invariably throughout the history of apprenticeship relations.

Moreover, by adding to the individual choices model a simplified "guild-like" mechanism,

we are able to describe the joint evolution of apprenticeship markets and guild regimes. As generations accumulate valuable information about their own trait, the demand for training tends to decrease. As long as providing training is profitable for current entrepreneurs, they will support a guild organization that guarantees the enforcement of quantitative limitations in an otherwise competitive market. But as soon as such an organization is no longer able to generate positive profits for its members, it is abandoned in favor of a fully competitive market, where entrepreneurs are indifferent towards providing training. The observed evolution of the apprenticeship market appears to be in accordance with the predictions of our model. The dimension of the market for apprenticeship positions shows a declining trend for most of the period under observation. The cost of education, represented by apprenticeship premiums, does not appear to have followed such a linear path: real premiums weakly declined until the early 18th century and then sparked up and remained roughly stable throughout the rest of the 1700s. We argue that the jump in premiums is related to the end of a "guild regime" of the economy and the beginning of a "no-guild regime" characterized by a competitive market for apprenticeship positions.

The paper is structured as follows. In the remainder of this Section we discuss the existing literature. After the historical notes contained in Section 1.2, in Sections 1.3 and 1.4 we build our model of educational and occupational choice. Section 1.4 also shows how a model like ours yields a realistic prediction about the diverse social origins of apprentices. In Section 1.5 we describe the dynamic phenomena implied by our model and we illustrate how these predictions fit the observed characteristics and some aspects of the evolution of apprenticeship in the Modern Age. Section 1.6 provides some comments and a few concluding remarks. Some derivations that would tamper the readability of the main body of the paper appear in Appendix A (Subsection 1.7.1). Appendix B (Subsection 1.7.2) provides a description of the data sources used in Sections 1.4 and 1.5.

1.1.2 Related Literature

The present work is inspired by several lines of research in economics and economic history and aims at providing a contribution to at least some of them.

Economic historians have long been showing a particular interest in the institution of modern-age apprenticeship, with a particular focus on England, both because of the larger amount of available evidence and because of the attention devoted to the study of the English economy in the centuries preceding and paving the way for the Industrial Revolution. The classic in this literature is the book by Dunlop and Denman (1912).

More recent comprehensive accounts include Lane (1996), Humphries (2003) and Epstein (2003). The book edited by De Munck et al. (2007) provides interesting parallels between apprenticeship institutions in several pre-industrial economies, including Japan. Earle (1989), Grassby (1995) and Schwarz (1992) also dedicate several pages to apprenticeship in their treatment of other aspects of modern-age English social and economic history. Several detailed studies have also appeared in recent decades. Smith (1973) and Schwarz (1987) provide information about characteristics and dimensions of the London apprenticeship market in the 17th and 18th century. Yarbrough (1980) and Ben Amos (1991) analyze the social origins and the post-training outcomes of Bristol apprentices during the 17th century. More recently, Hamilton (1995, 1996, 2000) has extensively studies the apprenticeship phenomenon in Canada. The re-evaluation of apprenticeship as a training institution with beneficial social effects started, to the best of our knowledge, with the work of Epstein (1998). Since then, research on apprenticeship has been inevitably linked with the study of modern-age guilds and the nature and effects of their interventions on the English economy. Several authors have pursued the traditional view that guilds were distortive, anti-market and essentially conservative institutions, whose abolition was of great benefit to the onset of the Industrial Revolution. Others have been suggesting that, whatever the intentions, guilds contributed to the creation of a fertile economic terrain for the advent of the industrial system. Ogilvie (2004) and Epstein (2008) can be taken as respective representatives. The interest in the guild-regime economy and its effects on the transition towards industrialism inspire most of the recent historical research on apprenticeship. Wallis (2008) contains a thorough treatment of the microeconomics of apprenticeship. Wallis et al. (2010) provided evidence on the age of entry into apprenticeship. Leunig et al. (2011) and Minns and Wallis (2012, 2013) provide data for and an analysis of several central aspects of the apprenticeship market in 18th century London. Minns and Wallis (2013), in particular, is of central importance for our present work.

Some authors have dealt more specifically with the economic rationale of modern-age apprenticeship, in order to make sense of several peculiar characteristics, such as long training periods, the presence of an upfront premium, the peculiar way apprentices were trained and the guild's and public authorities' attitude towards apprentices (all of which we shall discuss in what follows). Elbaum and Singh (1995) build a model of apprenticeship as an institution allowing the training of financially constrained agents and requiring the specific characteristics of length and premium in order to be economically viable for both masters and apprentices. This view is similar to that proposed by Malcomson et al. (2003) for contemporary professional training. Tadelis and Rangel (2001), on the other hand, model apprenticeship as a human capital accumulation device which also works as

a quality-signaling mechanism allowing young agents to reveal private information about their own selves and masters to discriminate workers along quality lines, thus overcoming a natural adverse selection issue. To the best of my knowledge, the only paper to have so far studied the aggregate, economy-wide role of apprenticeship is De la Croix et al. (2017). There the authors compare the rising of formal professional training in knowledge-intensive societies (such as Western Europe) to the traditional in-family or in-clan transmission of knowledge, which persisted elsewhere. Both this paper and theirs deal with the aggregate effects of apprenticeship. Still, while their aim is to assess the effects of such institution on long-run comparative development, the present paper is directed to highlighting a possible interpretation as of its economic rationale. In general, though, by methodology and aims, the present work is mostly related to the line of research represented by such papers. More recently, Ben Zeev et al. (2017) have provided empirical support to some of the arguments related to this paper, e.g. the responsiveness of apprenticeship institutions to technological and demand shocks: according to the authors, this may be of help in interpreting apprenticeship as one of the key institutions in the transition to a modern economy in 18th century Britain.

This paper also relates to work in economics that does not directly deal with apprenticeship or other historical institutions. As for the labor markets literature, the model that we develop bears shares some characteristics with Gonzalez and Shi (2010). In their paper, workers searching for job vacancies are characterized by an incomplete knowledge of an invariant yet unknown personal trait that determines their success in finding and keeping jobs. Similar to our agents who learn from apprenticeship, the workers of Gonzalez and Shi (2010) learn from their successes and failures in finding a job at their desired wage. Such events lead them to update their opinions about their own worth. Papageorgiou (2014) also develops a model of search, in which the source of learning is performance on the job. The underlying structure of the model is similar to ours and Papageorgiou's (2014) worker also sort themselves into categories depending on the intensity of their opinion about themselves: those who are nearer to certainty about their own abilities search for or remain into the occupation they see themselves to be more fit for; those who are most uncertain will remain into their profession as long as they don't have a chance to experiment the other. Both papers here described differ from us in that they also model wages endogenously in a labor-macroeconomics perspective. In our model, occupational compensations are exogenous,whereas the cost of learning is partially endogenous.

Finally, this work is deeply related to and inspired by the growing literature on the transmission of immaterial assets, such as values and beliefs, across generations of economic agents. This is a young but vast field of research to which we cannot do justice in such

a brief review. A general comprehensive review has been published by Bisin and Verdier (2011). One of the most relevant applications to historical phenomena is the paper by Doepke and Zilibotti (2008) on the effects of value transmission on the mechanism underlying the rise of capitalism. On a similar line, Doepke and Zilibotti (2014) studied the effects of cultural transmission on individual occupational choices and its consequences for growth. The way parental decisions are modeled here also owes to a related paper by the same authors (Doepke and Zilibotti, 2017). Guiso et al. (2008) and Tabellini (2008) both rely on the intergenerational transmission of beliefs which, in their case, are essentially beliefs on the distribution of trust types in a population of agents. An approach that is more similar to ours, in that agents learn and update beliefs about unknown features that may be seen as characteristics of their own, has been adopted by Fernández (2013) and Fogli and Veldkamp (2011) to analyze the participation of women to the labor market. To the best of my knowledge, these are no further instances of research in economics on the intergenerational transmission of beliefs concerning own unobservable characteristics. A secondary aim of the present work is therefore to re-evaluate the importance of learning about oneself as a conveyor of transformation in the cultural environment and its economic consequences.

1.2 Historical introduction

1.2.1 The Statute of Artificers

Modern-age apprenticeship consisted of a contract between a *master* (a craftsman, trader or, more in general, any citizen who was a recognized member of a professional guild) and a private citizen, the *apprentice*, or, alternatively, his/her parents or tutors. In its standard form, the contract obligated the apprentice to serve the master and, often, to the upfront payment of a monetary fee (the *apprenticeship premium*), in exchange for receiving training and basic sustenance (food, shelter and, sometimes, a small allowance for personal expenses) by the master for the whole period of training.

Apprenticeship was established as the main form of professional training in England by the Statute of Artificers, a law passed in 1563, at the onset of the Elizabethan period. The Statute harmonized at a national level a practice that had been developing throughout the "golden age" of the craft guilds, its origins probably dating back to the early 13th century. Due to the lack of formal recognition, the nature and forms of apprenticeship before 1563 remain largely unknown. The Statute required apprenticeship contracts to take the form of written *indentures* signed by the interested parties. The written form was

compulsory for the contract to have legal validity. Throughout the 16th and 17th century, the practical importance of holding a written contract grew, as apprentices could more easily resort to courts to settle litigation among them and their masters. From the 17th century onwards, the verifiability of apprenticeship contracts became fundamental as the successful completion of training guaranteed the apprenticed worker a number of political and economic rights, most importantly the chance to become a free citizen (*freeman*) of the town where he/she had served as apprentice.[1]

The Statute of 1563 also established age requirements to enter apprenticeship and a *minimum length* for the training period, which was set at 7 years for most professions.[2] The Statute also fixed a *maximum number* of apprentices that could serve each master at one time. This limit was usually set at 3 or 4 apprentices per master, depending on the master's trade. Both the minimum length and the maximum intake requirements are likely to have been the result of some lobbying activity by the guilds over the public authority. There is evidence that the maximum intake limit was sometimes waived upon in specific instances, possibly after special events who required the reconstruction of the craft community: in an episode reported by Lane (1996), workers from Eastern Europe and Russia were invited to become apprentices in several London crafts after the Great Fire of 1666. Overall, though, many agree that both limits were effectively enforced by the joint policing of guilds and public authorities, at least up to the early 18th century. Interestingly enough, even though the 17th century was a period of profound social transformation and the 18th century saw consistent migratory movements within the whole country, the prescriptions of the Statute of Artificers were never officially amended. The average age of completion of the training period thus remained set around 21 years for girls and 24 years for boys, which preceded by roughly 3 years the respective average ages of marriage (Wrigley et al., 1989).

1.2.2 Apprenticeship premiums

The Statute contained no indications about monetary transfers among the parties involved. As it happened for other private contracts at that time, the parties treated the payment of the premium as part of a "gentlemen's agreement": although the premium was economically essential, the parties did not feel the need to mention it in the writ-

[1] At that time, England retained a system of multiple levels of citizenship, according to which individual accomplishments were rewarded with higher degrees of citizenship.

[2] This long requirement is among the most interesting and most often analyzed features of English apprenticeship: in other European countries where a minimum length was fixed by the law, this was usually shorter than in England (De Munck et al., 2007).

ten document establishing the contract itself. In any case, it seems that the practice of apprentices paying a premium arose in the mid-17th century at the latest. Premiums started being recorded early in the 18th century, as the public authority imposed taxation on payments over a certain thresholds (1709).

When the payment of a monetary premium became an ever-present characteristic of training contracts, apprenticeship was soon transformed into a *profitable* activity for craftsmen, a secondary source of profit beyond the production of goods and services. The regulation and management of the apprenticeship market became extremely relevant for craft guilds during the 17th century, to the point that Epstein (1998,2008) has interpreted guilds as institutions essentially devoted to guaranteeing the possibility of a profitable exchange of apprenticeship positions. For modern-age craft guilds, maintaining apprenticeship premiums high appears to have been more important, as a way to increase their associates' profits, than direct intervention on the domestic market for goods and services.

Guilds did not interfere with the ability of their associates to privately establish the price of the apprenticeship positions they supplied.[3] But, whenever they could, guilds would strictly enforce the maximum intake limit established by the Statue of Artificers. Thus it seems reasonable to analyze the market for apprenticeship positions as one distorted by the presence of an exogenously enforced upper quantitative limit to supply. This is the view we chose to adopt in this paper.

Apprenticeship premiums were not small if compared with the average income of households: according to available data, average apprenticeship premiums varied, in the period under observation, between the level of a year's average working class wage and the level of year's worth of profits for an average craftsman. Still, they were not high enough to discourage parents of relatively modest origins from apprenticing their children. At the same time, they must have been high enough to make employing apprentices a profitable activity for craftsmen, by exceeding the implicit and explicit costs the master had to sustain when hiring an apprentice. In any case, we should not forget that all these are statements about averages. Boys and girls from modest families were likely to self-select into less fashionable crafts, whose premiums lay at the bottom of the distribution. According to Minns and Wallis (2013), few families would in fact pay more than some months worth of their income as an apprenticeship premium. Apprenticeship premiums were indeed subject to considerable variations in time and among trades (see, for instance, Minns and

[3]The view that premiums were not supervised by the guild and that their determination was left to private autonomy is partially supported by existing evidence of masters applying a sort of price discrimination based on the apprentice's family income and managing to exact larger premiums from apprentices with wealthier background.

Wallis, 2013). In this paper, we shall always use an average measure of apprenticeship premiums, thus disregarding the important differences among professions. We chose to do so in order to highlight general trends rather than the particular destinies of individual professions.

1.2.3 Training and completion

Contrary to what we may expect, modern-age masters did not exert much effort into training apprentices. Most sources (see, for instance Humphries, 2003, and Lane, 1996) agree that apprentices would commonly be assigned menial tasks, very often unrelated to the professions they were paying to learn, at least in the first years of training. As De Munck and Soly (2007) point out, the didactic skills of craftsmen were rarely put to work and apprentices underwent a process that is effectively described by the expression "learning on the shop floor". They might have been "learning by doing" in the last years of their training period, but before then they were most likely "learning by watching others doing". In the meantime they were expected to do the tidying of working rooms, the shopping for their master's family and to perform other domestic tasks. By the time apprentices began to receive wider legal protection, complaints that were carried to court often concerned a lack of actual learning opportunities.

Still, hiring an apprentice was far from costless for the master, who was responsible for the sustenance of his apprentices. It seems though that other costs, mostly legal or connected with the (very real) risk of the apprentice fleeing or stealing from the master's house, prevailed over board and lodging expenses. The presence of the guild mitigated the magnitude of such costs for the entrepreneur: the master could rely on the guild to offer legal assistance and some form of insurance against losses on apprenticeship contracts. These considerations will be reflected by our modeling choice in the regards of the structure of net apprenticeship benefits obtained by entrepreneurs.

A surprising characteristic of modern-age apprenticeship was the low rate of completion of the training period. Several sources agree that effective rate of completion ranged around 40%-50%. Schwarz (1987) performs his calculations about the number of apprentices in London on the assumption that 10% of the indentured apprentices would abandon the training program every year, resulting in a theoretical completion rate near to 48%. The data on apprenticeship in 17th-century Bristol displayed by Ben Amos (1991) show that around 31% of apprentices would eventually earn the status of freemen; becoming a freeman necessarily required the completion of an apprenticeship and, in turn, giving up freedom rights after completing an apprenticeship seems justifiable only in case of death

or migration to another town, thus suggesting that the average completion rate was in all likelihood not larger than 45%-50%. Grassby (1995) reports similar values for London. Even accounting for the fact that 15%-20% of apprentices may have died or have been convicted during the training period, these figures suggest that completion of training concerned no more than 65% of those who were alive and free by the time of the legal duration of the contract.

Thus many must have left the master's shop on a voluntary basis. Needless to say, those who did not complete their apprenticeship could not legally practice the trade they were trained for. Most of them would move back to their places and income class of origin, even though there is narrative evidence that apprenticeship dropouts would then introduce several technical improvements in traditional tasks that they had learned about while being apprenticed (see Ben Amos, 1991, and Lane, 1996). In this paper we argue that the large number of unfinished apprenticeships does not necessarily represent a failure of the educational institution. If the main function of apprenticeship was to let agents learn about unobserved characteristics, an incomplete training might well be a sign of an *effective learning* taking place, whereby the agent acquires the belief that a traditional occupation is in fact what is best for him/her and that carrying on the training would only result in the loss of potential earnings.

1.2.4 The economic function of apprenticeship

The minimum length and maximum intake limit contained in the Statute, as well as the presence of the upfront premium, have led many to see apprenticeship uniquely as one of the devices used by guilds to distort markets in favor of their members. As a professional training institution, the declared objective of apprenticeship was to transfer working knowledge and operational skills from experienced masters to young workers starting their career. But the long training period required by law appears hardly justifiable under this light. Even taking into account the fact that masters did not actively teach apprentices, the period required for the formation of a complete craftsman would have surely been shorter, at least for some professions: it would be surprising to imagine that it took young adults 7 years to master the practice of book-selling. Such considerations have sparked the idea that the institution of apprenticeship acted as a barrier against entry in the crafts and professions market. The prospect of having to pay a premium to enter a 7-years long period of bare subsistence within a master's shop would have discouraged many to apply for such positions. Because it was impossible to work as craftsmen or professionals where guilds controlled the market unless one had completed

an apprenticeship, discouraging mass entry into apprenticeship would have the effect of a limitation to the number of potential new guild associates. At the same time, the master could profit from hiring an apprentice both through the premium and as a source of cheap labor for the duration of the training period.

On a different line, other authors have argued that a lengthy training period was necessary to overcome a natural asymmetry in the provision of training. Apprentices could indeed learn the profession in much shorter time but, had they been allowed to leave as soon as they mastered the techniques, the master would have paid for their training and all the advantages would have been obtained by the next employer, who would hire the already-trained worker. By forcing apprentices to stay for a longer time span that it was necessary, and thus work for their masters in exchange for sustenance allowances, the regulation guaranteed the master an incentive to hire apprentices in the first place, through the promise of cheap labor. Thus apprenticeship, with its seemingly irrational and exploitative structure, allowed for training to take place and would be, at least in this respect, socially beneficial. The view that apprenticeship was essentially a means of contrasting *hold-up* phenomenon in labor relationships has been advanced by Wallis (2008) and characterizes the models of Elbaum and Singh (1995) and Malcomson et al. (2003).

These views carry some limitations that have been highlighted in the context of the recent re-evaluation of the role and economic function of craft guilds in the Modern Age by scholars such as S. Epstein (1998, 2008). In particular, it is not clear why a large fraction of apprentices came from families of craftsmen and professionals who could have let their descendant access their or their colleague's professions without forcing them to a lengthy and technically unjustifiable period of training. While the *intentions* of guilds were undoubtedly distortive, we must take into account other possible functions accomplished by apprenticeship that may possibly turn out to prevail, in the actual effects of this institution on the economy, over the pure barrier and hold-up effect.

One of the main forms of intervention of craft guilds on the market was related to quality control: guilds periodically inspected the produce of their associates and those who did not abide by the guild's standards were given fines and sanctions. Under such requirements, it was in the interest of guild members to hire high-quality workers and apprenticeship would provide workers with a "quality certificate" that could be profitably spent on the labor market. At the same time, this certification allowed the best workers to signal their quality by means of their activity as apprentices. This is, in a nutshell, the view proposed by Tadelis and Rangel (2001). The way of training described above, though, seems far from the best means of allowing quality signals to spring from young apprentices.

During the training period, youths were hardly tested. Moreover, the long duration of apprenticeships receives no satisfactory explanation within this framework.

Summarizing, the traditional view of apprenticeship as a barrier to entry and the hold-up theory account for the excessive length and the peculiar passive mode of training, but cannot satisfactorily explain the fact that craftsmen resorted to apprenticing their own descendants. The quality-signaling framework, which is compatible with the last fact, gives no solution to the issue of training length. In this paper, we advance the hypothesis that apprenticeship was valuable to agents as it allowed them to learn about some unobservable characteristic of their own. This explains why both workers and craftsmen normally apprenticed their children. Moreover, entrepreneurs and their guilds acted in a way to use apprenticeship as a source of safe secondary profits as long as this was possible. Hence our view does not disregard the distortive intentions of guilds, but assigns to apprenticeship a socially beneficial role that does not generate any inconsistency with the observed characteristics of this institution.

1.2.5 The decline of the guild regime

The guild-regime economy flourished until the early 18th century, when it began declining. Throughout the second half of the 17th century, the dimension of the apprenticeship market had been decreasing, as we shall show using evidence from the City of London. The (admittedly scarce) evidence about apprenticeship premiums in those years induces us to infer a slow decline in real premiums paid for training. Even though urban centers were subject to steady migratory inflows, the total demand for apprenticeship positions was reduced. This cannot be attributed to a decline in the crafts' vitality, as crafts and professional trades continued being the main entrepreneurial activities, especially in the City of London, throughout the 18th century. The new modes of industrial production were initially confined to Northern England and the first industrial plants were established around London only at the beginning of the 19th century. Nor were crafts being overwhelmed, at least until 1750-60, by the growing new sectors of overseas trade and financial services, which then became the leaders of the City's economic prosperity in the 19th century.

Kellett (1958) provides a thorough account of how guilds progressively lost the grip on aspects of the economy they had been successfully controlling only a few decades before. The phenomenon was certainly more complex than what the aims of this paper justify us to present. Guilds had long suffered from the illegal competitive pressure of "unfreemen" who practiced the trade at the outskirts of town, thus avoiding guild control and putting

downside pressure on prices by lowering the average quality of products on the market. Until the early 18th century, public authorities had backed the guilds in their fight against illegal traders. A shift in the authority's political orientation from mercantilism to economic liberalism induced town legislators to defend the guilds' prerogatives with weaker zeal since the late 17th century.[4]

Forced to devote more resources to fighting unlawful competition by its own means, early 18th-century guilds began transferring to individual members most of the apprenticeship enforcement costs. This, for instance, implied that individual craftsmen would have to pay for legal costs arising from lawsuits filed by apprentices and would have to personally take on the risk of apprentices misbehaving or disappearing (often to join the ranks of unlawful competitors). At the same time, and once again under the propulsive impetus of liberal ideas, national legislation turned to support the demands of apprentices for increased recognition of their rights. Hence, in a few years, entrepreneurs were burdened with the transfer of enforcement costs from guilds to themselves and with an increase in the magnitude of such costs due to the increased national attention for workers' rights.

According to Kellett (1958) the large amount of regulations and legislative interventions that appeared in the first 30 years of the 18th century reflect the tension between the guilds' attempt to preserve their powers and the public authority's intentions of paving the way for freer economic system without completely disrupting the guilds and the associated crafts. The Common Council's Act of 1712 can be taken as the starting point of the more intense phase in the decline of the guild regime. By 1750, few guilds retained the ability to effectively enforce the maximum intake limit. Apprenticeship contracts were being replaced by specialized labor contracts which did not guarantee any certainty of employment to the worker at the end of the contractual period, but which compensated skilled and semi-skilled workers with higher wages than the sustenance allowance given to apprentices. Traditional apprenticeship continued to exist throughout the second half of the 18th century, something which has been explained in other way than by recalling that apprenticeship was the only practical means of acquiring the status of freeman; the difficulty being that, with the fall of the guild regime, the residual advantages of such status are not clearly identified.

Later in this paper, we shall apply our idea of apprenticeship as a learning tool to provide for an explanation of apprenticeship persistence. Independently of such considerations, everyone agrees that by 1800 the dimension of the apprenticeship market had reached a physiological lower bound. Premium registration disappeared after 1780. The Statute

[4]Daniel Defoe, the English writer, was among the earliest and fiercest critics of the guild system (Lane, 1996). Adam Smith also had a poor opinion of the guild regime.

of Artificers was finally repealed in 1814, ratifying the end of the guild regime. Craft guilds turned into wealthy professional organizations with lobbying prerogatives but no actual powers on the markets for good and services. Since the early 19th century, the term "apprentices" was applied to poor children employed in industrial factories in exchange for food and shelter, which heavily contributed to inspire the 19th-century negative overview of the institution in its entirety.

1.3 The model: individual choices

In what follows we propose an intergenerational model of educational and occupational choice. Here we sum up its most peculiar features. First of all, the model has overlapping generations, in the sense that two generations (the "young" and the "old") co-exist in each period. Differently than traditional OLG models, the young agents do not make any choice, as they are subject to the educational choices of their parents. Still, the fact that they are the object of their parents' choices make their presence essential. Secondly, the modeled population is organized into dynasties and the model features a perfect genetic intergenerational transmission of a certain unobservable trait along dynastic lines. Agents in the model learn about this trait, but the latter is not affected by their actions. Thirdly, parents in our model are both altruistic (i.e. they take into account their children's welfare) and myopic (in essence, they anticipate their children's opinions but are unable to anticipate the fact that the environment their children will face is going to be different than the one they are currently facing).

The present Section illustrates in detail the assumptions of the model and describes its predictions in terms of individual educational and occupational choices. The following Section 1.4 is devoted to the analysis of the model's implications concerning aggregate outcomes.

1.3.1 Description of Timing and the Population

Time is discrete and we denote a generic period by $t \in \mathbb{N}_0$. At the beginning of each period $t \in \mathbb{N}_0$, a new *cohort* of agents is born. For reasons that will be made clear later on, the cohort of agents born at the beginning of period $t \in \mathbb{N}_0$ is denoted by C_{t+1} and called "cohort (or *generation*) $t+1$". Each cohort is composed of uncountably many agents and has the same, fixed measure, normalized to 1. We represent each cohort by means of a *countably infinite* set $\mathcal{I} \subset [0,1]$ endowed with a *non-atomic probability charge*

λ on $\left(\mathcal{I}, 2^{\mathcal{I}}\right)$.[5] For each $t \in \mathbb{N}_0$, each agent in cohort C_t is denoted by i_t, for some $i \in \mathcal{I}$. Each agent lives for two periods. Thus, for each $t \in \mathbb{N}_0$, agents belonging to cohort C_{t+1}, who are born at the beginning of period t, die at the end of period $t+1$ (agents in cohort C_0 are born in the non-modeled period -1 and die at the end of period 0). It follows that, in each period $t \in \mathbb{N}_0$, two cohorts overlap: cohort C_t, which is composed of "old" agents, and cohort C_{t+1}, which is composed of "young" agents. Because all cohorts have size 1, the total size of the population is 2 in each period.

For each $i \in \mathcal{I}$, the sequence of agents $(i_t)_{t \in \mathbb{N}_0}$ constitutes *dynasty* i. The first agent of dynasty i is i_0 and, for each $t \in \mathbb{N}_0$, i_{t+1} is the unique descendant of i_t. As we do for each single cohort, the set of all dynasties can also be identified with the set $\mathcal{I}$. Thus a set $A \subseteq \mathcal{I}$ can indicate a subset of dynasties or a subset of agents in a certain cohort and we shall specify the intended interpretation whenever confusion might arise. For instance, if $A \in 2^{\mathcal{I}}$ is the set defined by

$$A := \{i \in \mathcal{I} \,:\, i \text{ satisfies property } P_A\}$$

then $\lambda(A)$ can either denote the *measure* of all dynasties that satisfy property P_A or the *measure* of agents in a given cohort, say C_t for some $t \in \mathbb{N}_0$, that satisfy property P_A. Sometimes we shall call $\lambda(A)$ the *mass* or *size* of dynasties (or agents) that satisfy property P_A.

Each *dynasty* $i \in \mathcal{I}$ is assigned (by Nature) a *trait*, $\theta^i \in \{\ell, h\}$. The *trait assignment* function $\theta : \mathcal{I} \to \{\ell, h\}$ is defined by

$$\theta \,:\, i \mapsto \theta^i \tag{1.1}$$

so that the measure of dynasties with trait h is $\lambda\left(\theta^{-1}(\{h\})\right)$ and the measure of dynasties with trait ℓ is $\lambda\left(\theta^{-1}(\{l\})\right) = 1 - \lambda\left(\theta^{-1}(\{h\})\right)$. We assume that $\lambda\left(\theta^{-1}(\{h\})\right) \in (0,1)$: for each trait, there is a positive measure of dynasties with that trait.

Each agent who is a member of dynasty i shares trait θ^i with all her ancestors and descendants. In other words, such trait is *dynasty-specific* rather than individual-specific. The economic interpretation of traits will be given later. For now, it is important to notice that this trait is *not* a type in the sense first introduced by Harsanyi (1967), as different traits do not correspond to different information statuses of agents.

[5] See the rest of this Subsection for the motive of this choice.

Agents do *not* know their trait. For each $i \in \mathcal{I}$ and $t \in \mathbb{N}_0$, agent i_t of cohort C_t, *at the beginning of period* t, assigns to her dynasty's trait being h a probability $p^{i_t} \in (0,1)$ - and, obviously, a probability $1 - p^{i_t}$ to her trait being ℓ. This probability assignment is called agent i_t's *prior belief* about her trait. Notice that we exclude degenerate beliefs for simplicity.[6]

For all $t \in \mathbb{N}_0$, we define the *period* t *belief assignment* function $p_t : \mathcal{I} \to (0,1)$ by

$$p_t : i \mapsto p^{i_t} \tag{1.2}$$

so that, for all $Q \in \mathcal{B}_{(0,1)}$,

$$p_t^{-1}(Q) = \left\{ i \in \mathcal{I} : p^{i_t} \in Q \right\} \tag{1.3}$$

and we define, for all $t \in \mathbb{N}_0$, $\varphi_t : \{\ell, h\} \times \mathcal{B}_{(0,1)} \to [0,1]$ by letting, for all $\tau \in \{\ell, h\}$ and $Q \in \mathcal{B}_{(0,1)}$,

$$\varphi_t(\tau, Q) \coloneqq \lambda\left(\left\{ i \in \mathcal{I} : \theta^i = \tau \wedge p^{i_t} \in Q \right\}\right) = \lambda\left(\theta^{-1}(\{\tau\}) \cap p_t^{-1}(Q)\right) \tag{1.4}$$

Thus $\varphi_t(\tau, Q)$ is the mass of agents in cohort C_t who have trait τ and, at the beginning of period t, assign to having trait h a probability in Q.

Moreover, we assume that for all *Lebesgue-non-negligible* $Q \in \mathcal{B}_{(0,1)}$, $\varphi_0(\tau, Q) > 0$ for all $\tau \in \{\ell, h\}$. In other words, for each Lebesgue-non-negligible set of beliefs, there is a λ-non-negligible set of members of cohort C_0 of each trait with initial prior beliefs in Q. This implies, for instance, that there is a positive measure of agents whose trait is ℓ and assign a probability larger than .99 to having trait h. Similarly, there is a positive measure of agents whose trait is h and assign a probability smaller than .01 to having that trait.

Finally, for all $t \in \mathbb{N}_0$, we define $\gamma_t : \mathcal{B}_{(0,1)} \to [0,1]$ by letting, for all $Q \in \mathcal{B}_{(0,1)}$, $\gamma_t(Q) = \lambda\left(p_t^{-1}(Q)\right)$. Thus $\gamma_t(Q)$ is the measure of agents who assign to having trait h a probability in Q; in other words, γ_t is the marginal distribution of beliefs in the population at time t. Clearly, for all $t \in \mathbb{N}_0$ and $Q \in \mathcal{B}_{(0,1)}$,

$$\gamma_t(Q) = \varphi_t(\ell, Q) + \varphi_t(h, Q) \tag{1.5}$$

The reason for using a countable set of agents endowed with a charge instead of an uncountable set of agents, identifiable with the unit interval $[0,1]$, endowed with the usual Lebesgue measure, is that we wish to make use of a Law of Large Numbers on subsets of agents, in the spirit of Al-Najjar (2004). We refer to the Appendix for more details on this matter.

[6] The reason, of course, is that we intend to use Bayesian updating on these beliefs. And it is well-known that Bayesian updating has funny implications when events occur that were previously deemed impossible.

1.3.2 Lives

Consider now, for some $t \in \mathbb{N}_0$, agent i_t of cohort C_t. She is born at the beginning of period $t-1$. Agent i_t is *inactive* in period $t-1$, in the sense that she does not make any choice.[7] Instead, she is subject to the educational choice of her parent and, as such, she either becomes an apprentice (A) or she remains idle (NA). The parent's choice produces a random educational outcome, denoted by Y^{i_t} .

The *distribution* of Y^{i_t} depends on the parental educational choice. If the agent is idle, then she almost surely obtains no educational outcome, i.e.

$$Y^{i_t}(NA) = 0 \quad \text{w.p.} \quad 1 \tag{1.6}$$

If instead the agent becomes an apprentice, the random educational outcome is described by

$$Y^{i_t}(A) = \begin{cases} 1 & \text{w.p.} \quad \rho\mu_{\theta^i} \\ 0 & \text{w.p.} \quad 1-\rho \\ -1 & \text{w.p.} \quad \rho(1-\mu_{\theta^i}) \end{cases} \tag{1.7}$$

In (1.7), the educational outcome $Y^{i_t} = 0$ is a non-informative signal to the young agent. It is indeed the same outcome the agent would have obtained had she remained idle in period $t-1$, as one can see comparing (1.7) and (1.6). Both outcomes $Y^{i_t} = 1$ and $Y^{i_t} = -1$ are informative signals to the young agent. If $Y^{i_t} = 1$ we say agent i_t has been a successful apprentice. If $Y^{i_t} = 0$ we say agent i_t has been an unsuccessful or failed apprentice. Thus $\rho \in (0,1)$ is the probability of receiving an informative signal from education, conditional on being educated. Notice that ρ depends neither on time nor on the agent's identity. We take it as a time-invariant, exogenous measure of *apprenticeship effectiveness*.

Conditional on receiving an informative signal, the probability of agent i_t being a successful apprentice is $\mu_{\theta^i} \in (0,1)$. This probability only depends on the agent's dynastic trait: it does not depend on time nor on the agent's identity (that is, it depends on her identity only insofar as a specific trait is associated to her dynastic identity). For all $\tau \in \{\ell, h\}$, μ_τ is called *probability of success* associated with trait τ. We assume that $\mu_h \in \left(\frac{1}{2}, 1\right)$ and that $\mu_\ell = 1 - \mu_h \in \left(0, \frac{1}{2}\right)$, so that the probability that an agent with trait h succeeds equals the probability that an agent with trait ℓ fails, and vice versa. Moreover, we let $\Delta := \mu_h - \mu_\ell \in (0,1)$. Thus $\rho\Delta$ represents the *probabilistic advantage* of an agent with

[7] This is the reason why we prefer to depart from the traditional conventions and label as "cohort C_t" the set of agents born at the beginning of period $t-1$.

trait h over an agent with trait ℓ in succeeding as an apprentice.

At the beginning of period t, and after observing her educational outcome, agent i_t makes two choices: she chooses her *occupation* and she chooses the *education* of her son i_{t+1}. The occupational choice variable of agent i_t is $\omega^{i_t} \in \{W, E\}$. Here W is shorthand notation for the decision to become a worker in the "traditional sector" of the economy. E is shorthand notation for the decision to become an entrepreneur, which, in our contest, is equivalent to entering a *guilded craft*. The traditional sector referred to before is residual to the set of all guilded crafts. For simplicity, we shall name "workers" the agents who choose W and "entrepreneurs" or "craftsmen" the agents who choose E.

As it was the case for her parent, the educational choice variable of agent i_t is $a^{i_t} \in \{A, NA\}$. The choice set available to agent i_t for each of her two choices might be a restriction of, respectively, $\{W, E\}$ and $\{A, NA\}$: we shall say later when this might be the case.

Agent i_t's occupational choice leads to a random *occupational outcome*, denoted by X^{i_t}. The distribution of X^{i_t} depends on the agent's occupational choice. If the agent chooses to become a worker, she almost surely obtains an occupational outcome labeled 0, that is

$$X^{i_t}(W) = 0 \quad \text{w.p.} \quad 1 \tag{1.8}$$

If instead she chooses to become an entrepreneur, she may obtain two different occupational outcomes, distributed as follows:

$$X^{i_t}(E) = \begin{cases} 1 & \text{w.p.} \quad \mu_{\theta^i} \\ -1 & \text{w.p.} \quad 1 - \mu_{\theta^i} \end{cases} \tag{1.9}$$

We interpret $X^{i_t} = 1$ as an entrepreneurial success of agent i_t and $X^{i_t} = -1$ as an entrepreneurial failure of agent i_t. The labels are chosen in accordance with our notation for educational outcomes. Notice that the statement "$X^{i_t} = 0$" should be interpreted only as "agent i_t obtains the occupational outcome associated with working in the traditional sector": by no means we attach to it a qualitative assessment of success or failure. We wish to highlight the fact that both educational and entrepreneurial success depend on the agent's trait and, moreover, that the probability of being a successful entrepreneur (conditional on being an entrepreneur) and the probability of being a successful apprentice (conditional on becoming an apprentice and on apprentice being effective) are equal.

This consideration leads to a natural interpretation of dynastic traits as indicators of a certain set of characteristics that determine individual success in the crafts that constitute

the entrepreneurial occupations in the world described by this model. We choose to remain rather agnostic about the empirical meaning of dynastic traits. Several factors can be considered as drivers of entrepreneurial success, ranging from objective skills to specific human capital traits (De Nardi, 2004) to psychological and behavioral characteristics such as the degree of patience and/or risk aversion (see Doepke and Zilibotti, 2008, for instance).

Whatever our preferred interpretation of the empirical content of dynastic traits, the relevant assumptions here are that

- dynastic traits equally influence the probability of success of apprentices (conditionally on apprenticeship being effective) and of entrepreneurs;
- dynastic traits are *perfectly genetically* transmitted from parent to child.

The second assumption, in particular, deserves attention. By *perfect* transmission of the trait, we simply mean that, for all $t, s \in \mathbb{N}_0$ and all $i \in \mathcal{I}$, $\theta^{i_t} = \theta^{i_s} (= \theta^i)$. That is, as we said before, traits are dynasty-specific rather than individual-specific. By *genetic* transmission of the trait, we mean that transmission requires no active involvement of the parent or the child. We do not imply that the characteristics determining traits are actually part of the agents' genetic code. Skills must be taught to children; human capital investments must be planned and carried out by parents; the recent economic literature on preference and value transmission (see Bisin and Verdier, 2011, for a summary) convincingly claims that behavioral traits are at least semi-intentionally imposed by parents to their children. Moreover, if the transmission of such characteristics depends on parental actions, it is reasonable to think that such transmission would not be perfect, in the sense specified above. First of all, parents may (and usually do) consciously decide to impose on their children characteristics that differ from their own. And even if they attempted at inducing a perfect transmission of their own character, the outcome might be (and usually is) extremely noisy (see again Bisin and Verdier, 2011, or Doepke and Zilibotti, 2017). After all, in the same way a young agent, even conditional on her type, might turn out to be a good or bad apprentice with positive probability, the same young agent could be a good or bad learner and internalize parental teachings with different degrees of precision.

We intend the assumption of perfect genetic transmission to stand as a limiting case for a situation of imperfect intentional transmission in which traits of parents and children are highly correlated. The recent work by Clark (2014) supports the plausibility of this assumption: there is something beyond the plain transmission of wealth and assets that

determines the entrepreneurial success of industrial dynasties through centuries.

As dynastic traits in this model remain constant through time, they are subject to *intergenerational learning*. Recall that agents do not know their trait. Because traits affect the distribution of educational and occupational outcomes, agents are able to make inferences about their own trait simply by observing such outcomes. Because no outcome is exclusive of a specific trait (i.e. conditioning on the trait reduces but does not eliminate the randomness in outcomes), a finite number of observations do not allow agents to learn their trait with essential certainty. On the other hand, if what an agent learns can be at least partially transmitted to their descendants, information accumulates in time and agents of future generations can operate under a better understanding of their trait due to what past generations have learned. This is the spirit that informs the present work.

We impose, in this respect, an important restriction. Both educational and occupational outcomes can act as signals of dynastic traits to, respectively, agents who become apprentices (and whose apprenticeship is effective) and entrepreneurs. Yet *we assume that only what is learned from education* (that is, from apprenticeship) *can be transmitted from parents to children*. We do not claim that adult entrepreneurs do not learn from their entrepreneurial occupation outcomes. But we limit the possibility of transmitting to their children what they have been learning to the moment in which their children are born. After all, recall from Section 1.2 that the average age for the completion of an apprenticeship was around 24 years for men and 21 years for women and that the average age of marriage fluctuated around 27 years for men and 25 years for women. Having just completed their apprenticeship and yet to become established entrepreneurs (craftsmen were usually employed as journeymen for several years before being able to set up their own business), young parents could not have educated their children but in light of what they had learned so far. In our model, learning from one's occupational experience only happens at the end of the experience itself, by which time the son and daughters of entrepreneurs would already have left home and undertaken their own career.

Recall that $p^{i_{t-1}}$ is the probability assigned by agent i_{t-1} to having trait h, at the beginning of period i_{t-1}. At that time, agent i_t is born and inherits such belief from her parent. Thus, agent i_t inherits at birth the *trait-belief pair* $(\theta^i, p^{i_{t-1}})$. The belief p^{i_t}, held by agent i_t when she becomes active at the beginning of period t depends on her educational outcome. If agent i_t does not receive a meaningful signal from education (i.e. either she has not been apprenticed or her apprenticeship was ineffective) then she retains the belief inherited at birth, $p^{i_t} = p^{i_{t-1}}$. If, instead, agent i_t does receive a meaningful signal, she updates

her belief according to the evidence obtained, via Bayes' rule.[8] An educational success will lead her to increase the probability she attaches to having trait h; symmetrically, an educational failure will lead her to decrease such probability.

To be more precise, let us define the functions $q^+, q^- : [0,1] \to [0,1]$ by

$$q^+(p) := \frac{\mu_h p}{\mu_h p + \mu_\ell (1-p)} \quad ; \quad q^-(p) := \frac{\mu_\ell p}{\mu_\ell p + \mu_h (1-p)} \tag{1.10}$$

For any prior belief $p \in (0,1)$, $q^+(p)$ is the ratio between the probability of obtaining educational success ($\rho\mu_h$) times the probability of having trait h (p) and the unconditional probability of obtaining educational success ($\rho[\mu_h p + \mu_\ell (1-p)]$). Thus $q^+(p)$ is the Bayesian posterior probability assigned to having trait h, conditional on observing an educational success. Similarly $q^-(p)$ is the posterior probability assigned to having trait h, conditional on observing an educational failure.

It is easy to see from the definitions in (1.10) that $q^+ > \text{id}$ and q^+ is strictly increasing and concave on $(0,1)$; $q^- < \text{id}$ and q^- is strictly increasing and convex on $(0,1)$; finally, that $q^+ = (q^-)^{-1}$ and $q^- = (q^+)^{-1}$. The first fact indicates how a success induces the agent to update her belief upwards; how the ordering in two agents' belief is unchanged by updating if both are successful; and how the magnitude of such updating is smaller, the larger the initial belief: for an agent who is sufficiently confident of having trait h, a success is just a confirmation of her assessment and leads her closer to certainty, but not much closer than she already was. Similarly, the second fact denotes how a failure induces the agent to revise her belief downwards; how, as before, downward updating preserves the ordering of prior beliefs; and how the magnitude of updating is smaller, the smaller the initial belief. Finally, the third fact implies that, independently of the "starting point", two contrasting outcomes will lead you back to it. In particular, the order in which failure and successes occur over a certain number of trial does not matter: only the number of failure and successes determines the final posterior belief.

The relationship between $p^{i_{t-1}}$ and p^{i_t} is given by

$$p^{i_t}\left(Y^{i_t}, p^{i_{t-1}}\right) = \begin{cases} q^+\left(p^{i_{t-1}}\right) & \text{if} \quad Y^{i_t} = 1 \\ p^{i_{t-1}} & \text{if} \quad Y^{i_t} = 0 \\ q^-\left(p^{i_{t-1}}\right) & \text{if} \quad Y^{i_t} = -1 \end{cases} \tag{1.11}$$

Recall that Y^{i_t} is a random variable and that its distribution depends on $a^{i_{t-1}}$. Thus agent i_t's belief at the beginning of period t can be interpreted, from the point of view of an observer in period $t-1$, as a random variable whose distribution depends on $a^{i_{t-1}}$

[8] It is always possible to apply Bayes' rule because, for all $t \in \mathbb{N}_0$ and $i \in \mathcal{I}$, $p^{i_t} \in (0,1)$.

and $p^{i_{t-1}}$. Whenever we want to refer to this interpretation for beliefs that have not yet realized, we shall use the notation $P^{i_t}\left(a^{i_{t-1}}, p^{i_{t-1}}\right)$. Thus

$$P^{i_t}\left(NA, p^{i_{t-1}}\right) = p^{i_{t-1}} \quad \text{w.p.} \quad 1 \tag{1.12}$$

and

$$P^{i_t}\left(A, p^{i_{t-1}}\right) = \begin{cases} q^{+}\left(p^{i_{t-1}}\right) & \text{w.p.} \quad \rho\mu_{\theta^i} \\ p^{i_{t-1}} & \text{w.p.} \quad 1-\rho \\ q^{-}\left(p^{i_{t-1}}\right) & \text{w.p.} \quad \rho\left(1-\mu_{\theta^i}\right) \end{cases} \tag{1.13}$$

Now define a function $\mu : (0,1) \to (\mu_h, \mu_\ell)$ by

$$\mu(p) := p\mu_h + (1-p)\,\mu_\ell \tag{1.14}$$

This function associates to each possible belief $p \in (0,1)$ the expected (subjective) probability of entrepreneurial success of an agent holding belief p. It is easy to see that

$$\mu(p)\,q^{+}(p) + (1-\mu(p))\,q^{-}(p) = p \tag{1.15}$$

Using the result in (1.15),

$$\begin{aligned} \mathbb{E}_{p^{i_{t-1}}}\left[P^{i_t}\left(A, p^{i_{t-1}}\right)\right] &= (1-\rho)\,p^{i_{t-1}} + \rho\left\{\mu\left(p^{i_{t-1}}\right)q^{+}\left(p^{i_{t-1}}\right) + \left(1-\mu\left(p^{i_{t-1}}\right)\right)q^{-}\left(p^{i_{t-1}}\right)\right\} \\ &= (1-\rho)\,p^{i_{t-1}} + \rho p^{i_{t-1}} \\ &= p^{i_{t-1}} \end{aligned} \tag{1.16}$$

That is, the expected value of the apprenticed descendant's belief, conditional on the parent's belief and her decision to apprentice him, equals the parent's belief itself.

1.3.3 Individual choices

Fix now any period $t \in \mathbb{N}_0$. As we have seen, at the beginning of period t, agent i_t performs an occupational choice $\omega^{i_t} \in \{W, E\}$ and an educational choice $a^{i_t} \in \{A, NA\}$. In the period under analysis, the choice to apprentice one's child was legally unrestricted: every parent who could afford to pay the apprenticeship premium could decide to send her child to become an apprentice. Access to crafts, instead, was restricted to agents whose parents were entrepreneurs or who had completed an apprenticeship in their youth. In our model, we take this restriction into account. On the other hand, we impose no restriction on access to apprenticeship: even though some parents may have suffered from financial

constraints, the large number of apprenticeship of modest origins (see Section 1.4) testifies that their impact was not huge. In light of this evidence and in order to keep things simple, assume that no financial (liquidity or credit) constraints affect our agents.

The occupational choice set depends on the agent's dynastic and educational past. In particular the occupational options available to agent i_t are described by

$$O_{i_t}\left(\omega^{i_{t-1}}, a^{i_{t-1}}\right) = \begin{cases} \{W, E\} & \text{if} \quad (\omega^{i_{t-1}} = E) \vee (a^{i_{t-1}} = A) \\ \{W\} & \text{otherwise} \end{cases}$$

Agent i_t can choose to become an entrepreneur if either her parent i_{t-1} was an entrepreneur or if she has been apprenticed in period $t-1$. Otherwise - that is, if her parent was a worker *and* she has *not* been apprenticed in period $t-1$ - she can only become a worker. That is, she does not make any actual occupational decision. We assume that occupational choice is unrestricted for agents in cohort C_0 (whose youth and origins are not modeled), i.e. we assume that for all $i \in \mathcal{I}$, $O_{i_0} = \{W, E\}$.

We also assume, for the sake of simplicity, that all agents are *risk-neutral*. Agents value their present income, net of educational costs, and their child's gross income. In other words, agent i_t acts so to maximize a discounted sum of her own occupational payoff (minus the price of her son's apprenticeship if she decides to educate him) and her child's occupational income. Agent i_t does *not* take into consideration the fact that her child's income will be diminished by the costs her child will sustain for *his* descendant's education.

For all $t \in \mathbb{N}_0$, we can describe occupational payoffs in period t by a function $v_t : \{-1, 0, 1\} \to \mathbb{R}_+$ that maps occupational outcomes into (real) monetary prizes. Agents who become workers receive a fixed wage $w \in \mathbb{R}_{++}$ that remains constant in time. Because being a worker is associated with occupational outcome 0, for all $t \in \mathbb{N}_0$, $v_t(0) = w$. Becoming an entrepreneur, on the other hand, yields a random payoff. Within period t, randomness in this payoff is only due to the randomness in the occupational outcome (success or failure) associated with the entrepreneurial activity. Still, the payoffs corresponding to each of the two outcomes are *not* constant in time. In particular, the payoff obtained by entrepreneurs consists of a random component, that we simply call *"profit"*, that takes value $\pi \in \mathbb{R}_{++}$ or 0 according to whether the entrepreneur is (respectively) unsuccessful or successful. Each individual entrepreneur of any cohort is subject to such randomness in payoffs but, conditional on the occupational outcome, the amount of such payoff does not vary in time: a successful entrepreneur will obtain a profit π independently of her generation and, similarly, an unsuccessful entrepreneur will always obtain zero profit. *Independently* of occupational outcomes, each entrepreneur also obtains the

apprenticeship benefit b_t, that derives from her activity as an employer of apprentices. Such amount is awarded to all entrepreneurs and constitutes the non-random component of the entrepreneurial payoff, but it varies across periods. We shall dedicate a more extensive analysis to how b_t is determined in Section 1.4.

To sum up, the period t occupational payoff function is given by

$$v_t(X) = \begin{cases} w & \text{if} \quad X = 0 \\ \pi + b_t & \text{if} \quad X = 1 \\ b_t & \text{if} \quad X = -1 \end{cases} \tag{1.17}$$

Call c_t the apprenticeship premium in period t, which is the price that each agent in cohort C_t has to pay in order to educate her child.

The lifetime expected utility of agent i_t, as a function of her actions, can then be written as

$$U_{i_t}\left(\omega^{i_t}, a^{i_t}\right) = \mathbb{E}_{p^{i_t}}\left\{v_t\left(X^{i_t}\left(\omega^{i_t}\right)\right) - c_t \mathbf{1}_A\left(a^{i_t}\right) + \beta V_t^{i_{t+1}}\left(\omega^{i_t}, a^{i_t}, P^{i_{t+1}}\left(a^{i_t}, p^{i_t}\right)\right)\right\} \tag{1.18}$$

where, for all $q \in (0,1)$,

$$V_t^{i_{t+1}}\left(\omega^{i_t}, a^{i_t}, q\right) = \max_{\omega \in O_{i_{t+1}}\left(\omega^{i_t}, a^{i_t}\right)} \mathbb{E}_q\left[v_t\left(X(\omega)\right)\right] \tag{1.19}$$

Thus $V_t^{i_{t+1}}(\omega^{i_t}, a^{i_t}, P^{i_{t+1}}(a^{i_t}, p^{i_t}))$ is agent i_t's evaluation of agent i_{t+1}'s gross income, taking into account that i_{t+1} will optimally choose his occupation given his belief $P^{i_{t+1}}$, which is, in the eye of an observer in period t, a random variable whose distribution depends on i_t's educational choice a^{i_t} and i_t's current belief p^{i_t}, as explained in (1.13). Agent i_t's current occupational choice affects this quantity by potentially restricting the occupational choice set available to i_{t+1}.

We stress the fact that $V_t^{i_{t+1}}(\omega^{i_t}, a^{i_t}, q)$ is *not necessarily* the actual gross payoff that agent i_{t+1} will obtain from his occupational choice, but agent i_t's assessment of such value. This is due to the fact that agent i_t makes use of the payoff function v_t, which associates to entrepreneurial activity the current apprenticeship benefit b_t, not the apprenticeship benefit b_{t+1} that i_{t+1} will obtain if he decides to be an entrepreneur in period $t+1$.

Thus agent i_t is "myopic" in two ways. On the one hand, she does not take into account the costs that her child will sustain if he (the child) decides to educate his descendant. On the other hand, she does not consider that her child will most likely face different payoffs because of the variation in apprenticeship benefits.

In order to analyze the optimization performed by our agents, let us consider first the following problem

$$\max_{\omega \in \{W,E\}} \mathbb{E}_p \left[v_t \left(X \left(\omega \right) \right) \right] \tag{1.20}$$

As can be seen from (1.18), this is the problem an agent in cohort C_t with belief $p \in (0,1)$ would solve if $\beta = 0$, i.e. if she did not care about her child's welfare. It corresponds to an *isolated*, unrestricted occupational choice. Using (1.8), (1.9) and (1.17), problem (1.20) becomes

$$\max \left\{ w \, , \, \mathbb{E}_p \left[\mu_{\theta^i} \pi + b_t \right] \right\}$$

Using then the function μ introduced in (1.14), we rewrite it as

$$\max \left\{ w \, , \, \mu \left(p \right) \pi + b_t \right\}$$

We assume that, in case of indifference, every agent would choose to be a worker. Then there exists a unique solution to (1.20), given by

$$\hat{\omega} = W \iff p \leq \frac{w - \left(\mu_\ell \pi + b_t \right)}{\Delta \pi} \tag{1.21}$$

Define the function $\hat{p} : \mathbb{R}_+ \to \mathbb{R}$ by

$$\hat{p} \left(b \right) := \frac{w - \left(\mu_\ell \pi + b \right)}{\Delta \pi} \tag{1.22}$$

We call $\hat{p} \left(b \right)$ the *occupational threshold* corresponding to the apprenticeship benefit level b. From (1.21) we see that $\hat{p} \left(b \right)$ is the highest belief that induces an agent solving problem (1.20) to become a worker. As far as we know, it might be that $\hat{p} \left(b \right) \notin (0,1)$, meaning that either all agents prefer W to E or vice versa. We shall exclude the possibility that $\hat{p} \left(b \right) \leq 0$ later on. For now, just assume the following:

[A1] $\mu_\ell \pi < w < \mu_h \pi$

Assumption [A1] simply states that the profit expected by an entrepreneur who believes to have trait ℓ with essential certainty would be smaller than the working wage; similarly, the profit expected by an entrepreneur who believes to have trait h with essential certainty would be larger than the working wage. This assumption makes problem (1.20) non-trivial. An agent with trait ℓ who knew her trait should choose to be a worker rather than an entrepreneur *in absence of apprenticeship benefits*. Similarly, an agent with trait h who knew her trait should choose to be an entrepreneur *independently of apprenticeship benefits*. The second inequality in [A1] ensures that $\hat{p} \left(b \right) < 1$ for all $b \in \mathbb{R}_+$.

Let $\hat{\omega}^{i_t}$ be the choice that agent i_t would make if she solved problem (1.20). We call $\hat{\omega}^{i_t}$ agent i_t's *preferred occupation.* We should ask ourselves why and how can the occupation *actually* chosen by i_t differ from $\hat{\omega}^{i_t}$. This is certainly the case if $\hat{\omega}^{i_t} = E$ but agent i_t neither descends from an entrepreneur parent, nor has she been apprenticed. A less evident but equally plausible reason is that agent i_t's occupational choice be driven by her care for her descendant. Because apprenticeship has the same cost to workers and entrepreneurs and because the descendant of an entrepreneur is never restricted in his occupational choice, the last instance may occur only when $\hat{\omega}^{i_t} = W$ but i_t thinks that her child i_{t+1} will want to become an entrepreneur, and thus choosing to become an entrepreneur herself ($\omega^{i_t} = E$) will allow *him* to do so without the necessity of an apprenticeship. But no apprenticeship implies that i_{t+1} will hold the same belief as his parent which, as $\hat{\omega}^{i_t} = W$ implies, should induce him to choose occupation W. Thus agent i_t will always choose her preferred occupation, but for the case in which she is constrained to choose W. Even in this case, agent i_t's *occupational and educational choices are independent.*

Unconstrained working parents' educational choice

For the moment, let us neglect the possibility that agent i_t be constrained in her occupational choice and focus on her educational choice. Suppose i_t chooses to be a worker - which, since her occupational choice is unconstrained, is equivalent to saying that $p \leq \hat{p}(b_t)$. Then, for $a^{i_t} \in \{A, NA\}$,

$$\begin{aligned} U_{i_t}\left(W, a^{i_t}\right) &= \mathbb{E}_{p^{i_t}}\left\{w - c_t \mathbf{1}_A\left(a^{i_t}\right) + \beta V_t^{i_{t+1}}\left(W, a^{i_t}, P^{i_{t+1}}\left(a^{i_t}, p^{i_t}\right)\right)\right\} \\ &= w - c_t \mathbf{1}_A\left(a^{i_t}\right) + \beta \mathbb{E}_{p^{i_t}}\left[V_t^{i_{t+1}}\left(W, a^{i_t}, P^{i_{t+1}}\left(a^{i_t}, p^{i_t}\right)\right)\right] \end{aligned} \tag{1.23}$$

In particular, if $a^{i_t} = NA$, $P^{i_{t+1}}(a^{i_t}, p^{i_t}) = p^{i_t}$ and[9] i_{t+1} will be expected to become a worker, so that

$$U_{i_t}(W, NA) = w + \beta \mathbb{E}_{p^{i_t}}\left[V_t^{i_{t+1}}\left(W, a^{i_t}, P^{i_{t+1}}\left(a^{i_t}, p^{i_t}\right)\right)\right] = (1+\beta)\, w \tag{1.24}$$

If, instead, $a^{i_t} = A$, then using (1.13)

$$\begin{aligned} U_{i_t}(W, A) &= w - c_t + \beta\left\{(1-\rho)\, V_t^{i_{t+1}}\left(W, A, p^{i_t}\right) + \right. \\ &\left. +\rho\left[\mu\left(p^{i_t}\right) V_t^{i_{t+1}}\left(W, A, q^{+}\left(p^{i_t}\right)\right) + \left(1 - \mu\left(p^{i_t}\right)\right) V_t^{i_{t+1}}\left(W, A, q^{-}\left(p^{i_t}\right)\right)\right]\right\} \end{aligned} \tag{1.25}$$

Since $p^{i_t} \leq \hat{p}(b_t)$, one of the following two must hold:

[9] Here "and" has no consequential interpretation: even though a belief p^{i_t} would induce i_{t+1} to become a worker, we must recall that i_{t+1} will become a worker because his parent is a worker and he will not have been apprenticed.

1. $\hat{p}(b_t) \geq q^+(p^{i_t}) > p^{i_t} > q^-(p^{i_t})$: agent i_{t+1} is expected to choose W independently of his educational outcome;

2. $q^+(p^{i_t}) > \hat{p}(b_t) \geq p^{i_t} > q^-(p^{i_t})$: agent i_{t+1} is expected to choose W unless he obtains a positive educational outcome, in which case he chooses E.

In case (1),

$$\begin{aligned} U_{i_t}(W,A) &= w - c_t + \beta\left\{(1-\rho)w + \rho\left[\mu\left(p^{i_t}\right)w + \left(1-\mu\left(p^{i_t}\right)w\right)\right]\right\} \\ &= (1+\beta)w - c_t \\ (\forall c_t \geq 0) &\leq (1+\beta)w \\ &= U_{i_t}(W,NA) \end{aligned}$$

Before we assumed that, in case of indifference among the two occupations, every agent would choose to be a worker. Now we make a similar assumption and require that every agent who is indifferent between apprenticing and not apprenticing her descendant will opt for not apprenticing him. This implies that, if agent i_t has a belief p^{i_t} such that $q^+(p^{i_t}) \leq \hat{p}(b_t)$, she will *not* apprentice her child. After all, agent i_t is interested in her child's belief $p^{i_{t+1}}$ only insofar as the latter allows him to make a certain occupational choice. If the child is expected to choose W independently of his belief, then there is no incentive for agent i_t to invest in education.

In case (2), on the other hand,

$$U_{i_t}(W,A) = w - c_t + \beta\left\{(1-\rho)w + \rho\left[\mu\left(p^{i_t}\right)\left(\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t\right) + \left(1-\mu\left(p^{i_t}\right)\right)w\right]\right\}$$

which is equivalent to

$$U_{i_t}(W,A) = U_{i_t}(W,NA) - c_t + \beta\rho\mu\left(p^{i_t}\right)\left[\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t - w\right] \tag{1.26}$$

It is a matter of some algebra (relegated to the Appendix) to show that agent i_t will choose to educate her son if and only if

$$p^{i_t} > \frac{c_t + \beta\rho\mu_\ell\left[w - (\mu_\ell\pi + b_t)\right]}{\beta\rho\Delta(\pi + b_t - w)} \tag{1.27}$$

Define the function $\underline{p} : \mathbb{R}^2_+ \to \mathbb{R}$ by letting, for all $(b,c) \in \mathbb{R}^2_+$,

$$\underline{p}(b,c) := \frac{c + \beta\rho\mu_\ell\left[w - (\mu_\ell\pi + b)\right]}{\beta\rho\Delta(\pi + b - w)} \tag{1.28}$$

Thus by (1.27) $\underline{p}(b_t, c_t)$ is the largest value of beliefs that induces a working parent not to apprentice her child. We call this value *lower educational threshold.* Notice that

$$\underline{p}(b_t, 0) = q^-(\hat{p}(b_t)) \tag{1.29}$$

When apprenticeship is free, all agents whose children *might* become entrepreneurs if apprenticed decide to apprentice them. Education is free and the fact that working is always an option and that the working wage does not depend on beliefs constitute an *occupational insurance* mechanism in favor of the child.

On the other hand, a large apprenticeship premium might induce *no* working parent to educate their children. Indeed (as shown in the Appendix) if

$$c_t \geq \frac{\beta\rho\left[w - (\mu_\ell \pi + b_t)\right]\left[(\mu_h \pi + b_t) - w\right]}{\pi} \tag{1.30}$$

no apprenticeship will be chosen by working parents.

Entrepreneur parents' educational choice

We now move to the analysis of educational choices performed by entrepreneur parents. Agent i_t chooses to be an entrepreneur only if $p^{i_t} > \hat{p}(b_t)$. For $a^{i_t} \in \{A, NA\}$,

$$\begin{aligned} U_{i_t}\left(E, a^{i_t}\right) &= \mathbb{E}_{p^{i_t}}\left\{\mu_{\theta^i}\pi + b_t - c_t \mathbf{1}_A\left(a^{i_t}\right) + \beta V_t^{i_{t+1}}\left(W, a^{i_t}, P^{i_{t+1}}\left(a^{i_t}, p^{i_t}\right)\right)\right\} \\ &= \mu\left(p^{i_t}\right)\pi + b_t - c_t \mathbf{1}_A\left(a^{i_t}\right) + \beta \mathbb{E}_{p^{i_t}}\left[V_t^{i_{t+1}}\left(W, a^{i_t}, P^{i_{t+1}}\left(a^{i_t}, p^{i_t}\right)\right)\right] \end{aligned} \tag{1.31}$$

If $a^{i_t} = NA$, then $P^{i_{t+1}}(a^{i_t}, p^{i_t}) = p^{i_t}$ almost surely and i_{t+1} will choose to become an entrepreneur (he can, independently of the education received, because he descends from a parent who is currently an entrepreneur). Thus

$$U_{i_t}(E, NA) = (1+\beta)\left(\mu\left(p^{i_t}\right)\pi + b_t\right) \tag{1.32}$$

While, when $a^{i_t} = A$, agent i_{t+1} will hold an updated belief and, as seen before,

$$\begin{aligned} U_{i_t}(E, A) &= \mu\left(p^{i_t}\right)\pi + b_t - c_t + \beta\left\{(1-\rho)V_t^{i_{t+1}}\left(W, A, p^{i_t}\right) + \right. \\ &\left. +\rho\left[\mu\left(p^{i_t}\right)V_t^{i_{t+1}}\left(W, A, q^+\left(p^{i_t}\right)\right) + \left(1 - \mu\left(p^{i_t}\right)\right)V_t^{i_{t+1}}\left(W, A, q^-\left(p^{i_t}\right)\right)\right]\right\} \end{aligned} \tag{1.33}$$

Since $p^{i_t} > \hat{p}(b_t)$, one of the following two must hold:

1. $q^+(p^{i_t}) > p^{i_t} > q^-(p^{i_t}) > \hat{p}(b_t)$: agent i_{t+1} is expected to choose E independently of his educational outcome;

2. $q^{+}\left(p^{i_t}\right) > p^{i_t} > \hat{p}\left(b_t\right) \geq q^{-}\left(p^{i_t}\right)$: agent i_{t+1} is expected to choose E unless he obtains a negative educational outcome, in which case he chooses W.

We noticed before how a working parent would never apprentice her son if she expects him to be a worker independently of his educational outcome. We cannot be so fast in ruling out that a parent in situation (1) may apprentice her descendant. The reason is that while the expected occupational payoff of a worker does not depend on his belief, the payoff that the parent expects her entrepreneur-son to have *does* depend on the latter's belief $p^{i_{t+1}}$. Then an entrepreneur parent might want to apprentice her son with the only aim to make him more "optimistic" about his abilities as an entrepreneur. This, though, is not the case: to understand why, recall the result in (1.16) that

$$\mathbb{E}_{p^{i_t}}\left[P^{i_{t+1}}\left(A, p^{i_t}\right)\right] = p^{i_t}$$

Conditionally on the effectiveness of apprenticeship, the educating parent expects her child to be a successful apprentice with probability $\mu\left(p^{i_t}\right)$ and an unsuccessful one with probability $1-\mu\left(p^{i_t}\right)$. Because $\mu\left(p^{i_t}\right) q^{+}\left(p^{i_t}\right) + \left(1-\mu\left(p^{i_t}\right)\right) q^{-}\left(p^{i_t}\right) = p^{i_t}$, the expected value of the *updated belief* is, under the parent's evaluation, p^{i_t}. But since an ineffective apprenticeship directly imposes the parental belief upon the descendant, the unconditional expected value of i_{t+1}'s belief is, once again, p^{i_t}. It follows that the expected value of the descendant's entrepreneurial payoff, as evaluated by a parent holding belief p^{i_t}, is $\mu\left(p^{i_t}\right)\pi + b_t$, which is exactly the payoff the parent expects to obtain herself in the current period. Then apprenticing the descendant would be a waste of resources as the apprenticed descendant is expected to become an entrepreneur and obtain the same payoff as the non-apprenticed one.

We can now also say something about the relationship between b_t and w. Suppose it were the case that $b_t \geq w$. Then all agents in cohort C_t would become entrepreneurs since, even in the case of failure, the apprenticeship benefit obtained would be larger than the working wage. Accordingly, they will expect all of their descendants to become entrepreneurs, irrespective of their beliefs. Hence none of them would apprentice her descendant. The consequence is that no apprenticeship would exist in period t, so that $b_t = 0$, which contradicts our previous assumption which necessarily requires $b_t > 0$. Therefore, for all $t \in \mathbb{N}_0$, $b_t < w$.

In fact, we can claim by a similar reasoning that

$$\forall t \in \mathbb{N}_0\ ,\ b_t < w - \mu_\ell \pi \tag{1.34}$$

As for case (2), we can show (and do so in the Appendix) that

$$U_{i_t}\left(E, A\right) = U_{i_t}\left(E, NA\right) - c_t + \beta\rho\left[\mu\left(p^{i_t}\right)\left(\mu\left(q^{+}\left(p^{i_t}\right)\right)\pi + b_t - \pi - w\right) + w - b_t\right] \tag{1.35}$$

and that agent i_t will educate her son if (recall that $b_t < w$)

$$p^{i_t} > \frac{-c_t + \beta\rho\mu_h \left[w - (\mu_\ell \pi + b_t)\right]}{\beta\rho\Delta \left(w - b_t\right)} \tag{1.36}$$

Similarly to what we did for the working parent, define the function $\overline{p} : \mathbb{R}^2_+ \to \mathbb{R}$ by letting, for all $(b, c) \in \mathbb{R}^2_+$,

$$\overline{p}\left(b, c\right) := \frac{-c + \beta\rho\mu_h \left[w - (\mu_\ell \pi + b)\right]}{\beta\rho\Delta \left(w - b\right)} \tag{1.37}$$

By (1.36) $\underline{p}\left(b_t, c_t\right)$ is the smallest value of beliefs that induces an entrepreneur parent not to apprentice her child. We call this value *upper educational threshold*. Notice that

$$\underline{p}\left(b_t, 0\right) = q^+\left(\hat{p}\left(b_t\right)\right) \tag{1.38}$$

Compare this with (1.29). There we interpreted this condition as representative of the fact that, for parents with sufficiently high beliefs, apprenticeship might induce the descendant to become an entrepreneur which would be optimal in the case of educational success; so that free apprenticeship implied that every parent whose son had *a chance* to become an entrepreneur, would put him in the condition to do so. It might seem surprising at first that every entrepreneur parent whose son has a chance to become *a worker* would apprentice him. The underlying mechanism here relies on the fact that being an entrepreneur is not optimal for an agent who has trait ℓ. If agent i_t is an entrepreneur but places high enough probability on her trait being ℓ, she evaluates that her child might profit from being educated, because education is informative (albeit imperfectly) and may induce the child to work rather than become an entrepreneur, were a signal of him being trait ℓ received as a result of the educational experience.

As before, a high enough apprenticeship premium may induce *no* entrepreneur to apprentice their descendants. This happens when

$$c_t \geq \frac{\beta\rho \left[w - (\mu_\ell \pi + b_t)\right] \left[(\mu_h \pi + b_t) - w\right]}{\pi} \tag{1.39}$$

Condition (1.39) just obtained is the same as (1.30): the apprenticeship premium influences the existence of a domain of beliefs where some agents educate their children equally for current workers and current entrepreneurs.

Constrained working parents

We must now deal with the possibility that some agents are restricted in their occupational choice. Before we assumed that members of cohort C_0 were free to choose their preferred

occupation. This is enough to understand that the situation under analysis can only arise because of the evolution of the occupational threshold in time.

For all $t \in \mathbb{N}_0$, we denote by $\underline{p}_t, \hat{p}_t, \overline{p}_t$ the actual educational and occupational thresholds in period t. This is short-hand notation for, respectively, $\underline{p}(b_t, c_t)$, $\hat{p}(b_t)$ and $\overline{p}(b_t, c_t)$, where b_t and c_t denote the *actual* levels of apprenticeship benefits and apprenticeship premium determined in period t.

Suppose for a moment that, for all $s \in \{1, \ldots, t\}$, $\hat{p}_s = \hat{p} \in (0, 1)$; that is, suppose the occupational thresholds has not changed up to time t. Suppose agent i_t is restricted in her occupational choice, in the sense that $p^{i_t} > \hat{p}$ but $O_{i_t} = \{W\}$. Then it is necessary that $\omega^{i_{t-1}} = W$, for otherwise i_t could access entrepreneurship because of her parent's occupation. But it is equally necessary that i_t has *not* been apprenticed, which implies that $p^{i_{t-1}} = p^{i_t} > \hat{p}$. Then we find that, in period $t-1$, i_{t-1} has made a restricted occupational choice just like his daughter is about to in period t. Iterating this reasoning backwards, we obtain $p^{i_0} > \hat{p}$ and $\omega^{i_0} = W$, which is absurd since i_0's choice was not restricted.

Then, if in period t agent i_t is restricted in her occupational choice, there must have been some ancestor of hers, say i_s with $s \in \{1, \ldots, t-1\}$, who was the first member of dynasty i to be restricted in his occupational choice. This implies that i_{s-1} was a worker *by his own choice*, i.e. $p^{i_{s-1}} \leq \hat{p}_{s-1}$, while the fact that i_s is restricted implies $p^{i_s} > \hat{p}_s$. But i_s cannot have been apprenticed, for otherwise he would not be restricted. Thus $p^{i_s} = p^{i_{s-1}}$, which implies $\hat{p}_{s-1} > \hat{p}_s$. Thus, a *necessary* condition for *some* agent to be restricted in *some* period $t \in \mathbb{N}_0$ is that there is $s \in \{1, \ldots, t\}$ such that $\hat{p}_{s-1} > \hat{p}_s$.

As one can easily see, this condition is far from being sufficient. We notice indeed that if, say, $\hat{p}_s < \hat{p}_{s-1}$ but $\hat{p}_s \geq \underline{p}_{s-1}$, then even though there are sons of workers who end up wanting to become entrepreneurs while retaining the parental belief (i.e. all the sons of agents in C_{t-1} with belief in $(\hat{p}_s, \hat{p}_{s-1}]$), *none* of those is constrained, as all of them are *apprenticed* sons of workers. All of them, in other words, will want and will become entrepreneurs in period s.

Suppose agent i_t is constrained to choose W but $p^{i_t} > \hat{p}(b_t)$ and let us examine her educational choice. Let us denote by $U_{i_t}(W^c, a^{i_t})$ the lifetime utility of such agent, in order to highlight the fact that W was the outcome of a constrained choice. If she does not apprentice her child, then

$$U_{i_t}(W^c, NA) = (1 + \beta) w$$

as before. The expected lifetime utility to agent i_t if she apprentices her son is

$$\begin{aligned} U_{i_t}(W^c, A) &= w - c_t + \beta \left\{ (1-\rho) \left(\mu \left(p^{i_t} \right) \pi + b_t \right) + \right. \\ &+ \left. \rho \left[\mu \left(p^{i_t} \right) \left(\mu \left(q^{+} \left(p^{i_t} \right) \right) \pi + b_t \right) + \left(1 - \mu \left(p^{i_t} \right) \right) w \right] \right\} \end{aligned}$$

By comparing this to (1.26) one immediately notices that, whenever there are unconstrained workers that apprentice their sons, all *constrained* workers will apprentice their sons. Because the hypothesis will always be satisfied in this model, this shows that *no two consecutive members* of the same dynasty can ever be constrained workers. We have thus shown that a necessary and sufficient condition for the existence of constrained workers in period $t \in \mathbb{N}$ is that $\hat{p}_t < \underline{p}_{t-1}$. The definitive solution to the issue of constrained worker will be given in Section 1.5, as it requires an analysis of the aggregate consequences of our model.

1.4 The model: aggregation

This Section and the next deal with the analysis of the aggregate implications of our model. The individual educational and occupational choices described in Section 1.3 are aggregated in order to obtain a stylized model of the market for apprenticeship positions, in which parents willing to educate their descendants constitute the demand side and entrepreneurs willing to employ apprentices constitute the supply side. Aggregation is performed separately for the cases of a market characterized by the presence of a craft guild and fully competitive markets. In the first case we say that the market works *under a guild regime*; in the second case, we shall talk about a *no-guild regime*. Consistently with the historical evidence discussed in Section 1.2, the guild regime is characterized by the enforcement of a maximum intake limit and by a lower level of apprenticeship-related costs accruing to individual entrepreneurs than those which prevail in a no-guild regime. In this paper, we do not provide a model of why guilds exist and, for them, we do not apply a principle of "institutions as equilibria": to keeps matter simple, we assume that the guild regime arises whenever the presence of the guild leads to positive profits for entrepreneurs on the apprenticeship market. Otherwise, the market works in a no-guild regime.

1.4.1 Occupational and educational pools

For each $t \in \mathbb{N}_0$, the occupational and educational choices described in the previous section determine the measure of "old" agents (members of cohort C_t) who become workers and

entrepreneurs, as well as the measure of young agents (members of cohort C_{t+1}) who become apprentices in period $t \in \mathbb{N}_0$. Let us, for the moment, neglect the possibility of constrained workers existing at time $t \in \mathbb{N}_0$. Under this assumption, and because of the absence of liquidity constraints, occupational and educational choices only depend on the agents' prior beliefs at the beginning of period t. We can then define correspondences that map apprenticeship premiums and benefits into subsets of $(0, 1)$ denoting the beliefs that will induce certain occupational or educational choices.

Because occupational choices only depend on apprenticeship benefits and not (at least, not directly) on the apprenticeship premium, we define the maps $\mathcal{W}, \mathcal{E} : \mathbb{R}_+ \to \mathcal{B}_{(0,1)}$ that associate to each level of apprenticeship benefits the sets of beliefs

$$\mathcal{W}(b) := (0, \hat{p}(b)] \tag{1.40}$$

and

$$\mathcal{E}(b) := (\hat{p}(b), 1) \tag{1.41}$$

which are, respectively, the set of beliefs that induce agents to become workers and the set of beliefs that induce agents to become entrepreneurs. For all $b \in \mathbb{R}_+$, such two sets are non-empty because $\hat{p}(b) \in (0, 1)$ (as explained in Section 1.3). We call $\mathcal{W}(b)$ and $\mathcal{E}(b)$, respectively, the *working pool* and the *entrepreneurship pool*. Together, they will be referred to as the *occupational pools*. Notice, though, that both are sets of beliefs, not sets of agents. The sets of agents that, in period t, will become workers and entrepreneurs are $p_t^{-1}(\mathcal{W}(b_t))$ and $p_t^{-1}(\mathcal{E}(b_t))$, their measures being $\gamma_t(\mathcal{W}(b_t))$ and $\gamma_t(\mathcal{E}(b_t))$, respectively.

Educational choices depend, on the other hand, on apprenticeship benefits and apprenticeship premiums. Then we define the map $\mathcal{A} : \mathbb{R}_+^2 \to \mathcal{B}_{(0,1)}$ that associates to each level of apprenticeship benefits and apprenticeship premium the set of beliefs

$$\mathcal{A}(b, c) := \left(\underline{p}(b, c), \overline{p}(b, c)\right) \tag{1.42}$$

Notice that $\mathcal{A}(b, c)$ can be the empty set (see (1.30) and (1.39)). We call $\mathcal{A}(b, c)$ the *apprenticeship pool*: it is the set of beliefs that induce agents to educate their children through apprenticeship. The set of agents who will apprentice their children in period t is $p_t^{-1}(\mathcal{A}(b_t, c_t))$, its measure is $\gamma_t(\mathcal{A}(b_t, c_t))$. As in the previous section, we shall often refer to the *actual* occupational and apprenticeship pools in period t by $\mathcal{W}_t$, $\mathcal{E}_t$ and $\mathcal{A}_t$.

1.4.2 Demand and supply of apprenticeship positions

Each parent that decides to educate her children contributes by one "unit" to the *aggregate demand for apprenticeship positions*, that constitutes the demand side of the *apprenticeship market*. Notice that the good exchanged on this market is not the apprentices' labor,

but apprenticeship positions. Hence those willing to become apprentices constitute the demand side of the market and those willing to hire apprentices the supply side. As agents in this model are λ-negligible, the initial assertion should be interpreted as stating that, for each $t \in \mathbb{N}$ and for all $(b, c) \in \mathbb{R}_+^2$, the demand for apprenticeship positions at (b, c) equals the measure of agents with beliefs belonging to the apprenticeship pool, i.e. $\gamma_t(\mathcal{A}(b, c))$. We can then define the *demand function* at period $t \in \mathbb{N}_0$ as $D_t : \mathbb{R}_+^2 \to [0, 1]$ by letting, for all $(b, c) \in \mathbb{R}_+^2$,

$$D_t(b, c) := \gamma_t(\mathcal{A}(b, c)) \tag{1.43}$$

The identification of the supply curve of the apprenticeship market is slightly more complicated. All entrepreneurs can hire apprentices. When the market operates under a *guild regime*, the maximum number of apprentices that each entrepreneur can hire is set to $m_{\max} \in \mathbb{R}_+$ by the guild. When the market operates under a *no-guild regime*, this limit does not exist (or is not enforced). We can describe this situation by letting $m_{\max} = +\infty$. The payoff each entrepreneur receives from his/her activity as an employer of apprentices constitutes the individual level of apprenticeship benefits, b. Apprenticeship benefits depend on three factors

- the apprenticeship premium $c \in \mathbb{R}_+$ paid by each apprentice to the employer-entrepreneur;
- the cost for an employer-entrepreneur of hiring each apprentice, $\phi \in \mathbb{R}_{++}$;
- the number $m \in \mathbb{N}$ of apprentices that an employer-entrepreneur hires.

The function that describes the dependence of apprenticeship benefits on these components is $b : \mathbb{R}_+ \times \mathbb{R}_{++} \times \overline{\mathbb{R}}_+ \to \overline{\mathbb{R}}_+$ defined by

$$b(c, \phi, m) := (c - \phi)\, m \tag{1.44}$$

As we discussed in Section 1.2, historical evidence points to a particular training mode that can be described as "learning on the shop floor". The master craftsman who employed apprentices exerted no direct effort in teaching them the professional skills they were paying to learn. The apprentice would mainly observe other workers in the shop and retain what he/she could. Hence it is reasonable to assume that the costs arising for the entrepreneur when hiring an apprentice do not depend on how many other apprentices are currently at work in the shop. Such costs, as we illustrated before, were mainly connected with the possibility of legal issues and apprentices' misbehavior. This is the

reason why we assume that the total costs from hiring m apprentices are linear and equal to ϕm. At the same time, in this model all apprentices pay the same premium c, which has to be interpreted as the average premium on the apprenticeship market. Thus the overall apprenticeship benefits are assumed to be described by (1.44). Moreover, as we have seen, the presence of the guild contributes to making the entrepreneur's individual burden lighter. For this reason we assume that the level of costs under the no-guild regime is higher ($\overline{\phi}$) than under the guild regime ($\underline{\phi}$).

Given the apprenticeship benefits functional form in (1.44), each entrepreneur then solves the following problem:

$$\mathcal{P}\left(c,\phi,m_{\max}\right) : \begin{cases} \max_{m\in\overline{\mathbb{R}}_+} & b\left(c,\phi,m\right) \\ \text{s.t.} & m \leq m_{\max} \end{cases} \tag{1.45}$$

Let the solution *correspondence* be denoted by $m^* : \mathbb{R}_+ \times \mathbb{R}_{++} \times \mathbb{R} \to \overline{\mathbb{R}}_+$. Under the no-guild regime, $m_{\max} = +\infty$ (so problem $\mathcal{P}$ is effectively unconstrained) and $\phi = \overline{\phi} \in \mathbb{R}_{++}$. The solution to $\mathcal{P}\left(c,\overline{\phi},+\infty\right)$ is easily seen to be

$$m^{NG}\left(c\right) := m^*\left(c,\overline{\phi},+\infty\right) = \begin{cases} \{0\} & \text{if} \quad c < \overline{\phi} \\ \overline{\mathbb{R}}_+ & \text{if} \quad c = \overline{\phi} \\ \{+\infty\} & \text{if} \quad c > \overline{\phi} \end{cases} \tag{1.46}$$

Under the guild regime, $m_{\max} \in \mathbb{R}_+$ and $\phi = \underline{\phi} \in \mathbb{R}_{++}$. The solution to $\mathcal{P}\left(c,\underline{\phi},m_{\max}\right)$ is

$$m^{G}\left(c\right) := m^*\left(c,\underline{\phi},m_{\max}\right) = \begin{cases} \{0\} & \text{if} \quad c < \underline{\phi} \\ [0,m_{\max}] & \text{if} \quad c = \underline{\phi} \\ \{m_{\max}\} & \text{if} \quad c > \underline{\phi} \end{cases} \tag{1.47}$$

Notice that such solutions do not depend on the employer-entrepreneur's identity or her belief. Therefore all entrepreneurs will make the same hiring choice under a given apprenticeship market regime. This will make the supply for apprentices positions ultimately depend only on the market regime and the total measure of entrepreneurs and their common hiring choice for each level of the apprenticeship premium.

Let $b^{NG}\left(c\right) := b\left(c,\overline{\phi},m^{NG}\left(c\right)\right)$ and $b^{G}\left(c\right) := b\left(c,\underline{\phi},m^{G}\left(c\right)\right)$ be the levels of apprenticeship benefits corresponding to the apprenticeship premium c under the two market regimes. The measure of entrepreneurs active in period t, under the two regimes, is then given, for $R \in \{NG, G\}$ and $c \in \mathbb{R}_+$, by

$$e_t^R\left(c\right) := \gamma_t\left(\mathcal{E}\left(b^R\left(c\right)\right)\right) \tag{1.48}$$

As each entrepreneur supplies to the market $m^R(c)$ apprenticeship positions under regime R, the supply correspondence is given, for all $t \in \mathbb{N}_0$, $R \in \{NG, G\}$ and $c \in \mathbb{R}_+$, by

$$S_t^R(c) = m^R(c)\, e_t^R(c) = m^R(c)\, \gamma_t \left(\mathcal{E}\left(b^R(c)\right)\right) \tag{1.49}$$

In particular, under the no-guild regime,

$$S_t^{NG}(c) = \begin{cases} \{0\} & \text{if} \quad c < \overline{\phi} \\ \overline{\mathbb{R}}_+ & \text{if} \quad c = \overline{\phi} \\ \{+\infty\} & \text{if} \quad c > \overline{\phi} \end{cases} \tag{1.50}$$

while under the guild regime

$$S_t^G(c) = \begin{cases} \{0\} & \text{if} \quad c < \underline{\phi} \\ \left[0, m_{\max} e_t^G\left(\underline{\phi}\right)\right] & \text{if} \quad c = \underline{\phi} \\ \left\{m_{\max} e_t^G(c)\right\} & \text{if} \quad c > \underline{\phi} \end{cases} \tag{1.51}$$

Upon defining the apprenticeship benefits for the two market regimes, we can also re-express the aggregate demand as a function of the apprenticeship premium only. Thus we define, for all $t \in \mathbb{N}_0$ and $R \in \{NG, G\}$, the function $D_t^R : \mathbb{R}_+ \to [0,1]$ by letting, for all $c \in \mathbb{R}_+$,

$$D_t^R(c) := D_t\left(b^R(c), c\right) \tag{1.52}$$

1.4.3 Social Origins

At this stage, we can already comment on one prediction provided by our model. As long as apprenticeship premiums are smaller than the bound given by (1.30) and (1.39), the apprenticeship pool is non-empty and it contains the occupational threshold. That is, if b_t and c_t are such that $\mathcal{A}(b_t, c_t) \neq \emptyset$, then $\hat{p}(b_t) \in \mathcal{A}(b_t, c_t)$.

Recall now that we assumed in Section 1.3 that, for all non-negligible $Q \in \mathcal{B}_{(0,1)}$ and all $\tau \in \{\ell, h\}$, $\varphi_0(\tau, Q) > 0$. Suppose now that this property holds for some $t \in \mathbb{N}_0$ and choose any non-negligible $Q \in \mathcal{B}_{(0,1)}$ and $\tau \in \{\ell, h\}$. The set Q may be chosen so that all agents in $p_t^{-1}(Q)$ apprentice their descendant or that only some of them do or, again, that none of them does. In all cases, at least a fraction $(1-\rho)$ of those agents who have belief in Q at the beginning of period t will have descendants retaining the parental belief in period $t+1$. That is, however Q is chosen, it holds $\varphi_{t+1}(\tau, Q) \geq (1-\rho)\, \varphi_t(\tau, Q)$. But this implies that $\varphi_{t+1}(\tau, Q) > 0$. An inductive reasoning shows that $\varphi_t(\tau, Q) > 0$ holds for all $t \in \mathbb{N}_0$, however τ and Q are chosen.

Because of this conclusion, whenever the apprenticeship pool is not the empty set, there will be a positive measure of agents with beliefs in the apprenticeship pool and, in particular, in the intersections of the apprenticeship pool with the working pool and the entrepreneurship pool. This implies that, whenever someone is apprenticed (i.e. as long as the apprenticeship market is active), there is a positive measure of apprentices whose parents come from each occupational pool. This prediction of our model is well reflected by observations about modern-age apprenticeship.

Not only in London, but in most towns for which data are available, the social background diversity of apprentices appears to have been a constant characteristic throughout the 16th, 17th and 18th centuries. According to the observations collected by several authors, the percentage of apprentices whose parents worked within the guilded crafts range between 35% and 60%, with no substantial differences between the 16th and 18th centuries. The variability of this indicator mostly depends on whether liberal professions (legal, medical, etc.) are included in the account. The data displayed by Ben Amos (1991) for the town of Bristol in 1600-1630 report that around 40% of apprentices originated from families of craftsmen. Around 20% had parents who were employed in non-craft trades and liberal profession. The remaining 40% were descendants of husbandmen, urban laborers, members of the country gentry and sailors, all of which would fall under our "traditional occupation" comprehensive label. Similar figures are provided by Yarbrough (1980) in reference to 16th century Bristol, just after the passing of the Statute of Artificers. Minns and Wallis (2012) find that in late 17th century London, sons and daughters of craftsmen constituted 36% of the apprenticed population. Professions accounted for 27% and the remaining 37% originated from traditional backgrounds. Around the same time, the percentage of craftsmen's descendant in Bristol had almost reached 50%, possibly because London had become the main attractor for people from rural Southern England at its expenses. For 18th century London, Minns and Wallis (2013) report once again a percentage of 36% descendants of craftsmen and 22% descendants of merchants and other professionals, which leaves 41% descending from workers of the primary sector, laborers and country gentlemen. Table 4 in Minns and Wallis (2013) also shows how country gentlemen mostly entered apprenticeships in order to practice a liberal profession rather than a technical craft. Finally, Grassby (1995) reports similar figures for the city of London and the towns of Bristol and Southampton in the 17th century.

The indications of the above data are clear: throughout the period when apprenticeship was a mass educational phenomenon, it collected descendant of "traditionally occupied" families as well as sons and daughters of those that we labeled "entrepreneurs". The high and persistent degree of social diversity that characterized the apprenticing population

constitutes an important aspect of the high degree of social mobility that characterized modern-age England. Moreover, the large number of descendants of craftsmen and other professionals casts some doubts about whether apprenticeship worked only as a pass for guilded professions, since most of those youths could have accessed such occupations without the need of a lengthy and technically largely unnecessary training (see the discussion in Section 1.2).

1.5 The model: equilibria and dynamics

1.5.1 The apprenticeship market under the no-guild regime

Under the no-guild regime, the apprenticeship market works in a rather simple way. For any $t \in \mathbb{N}_0$, if $c > \overline{\phi}$, then $S_t^{NG}(c) = \{+\infty\}$: since the demand for apprenticeship positions is bound by the adult population's size, the competition among employer-entrepreneurs to hire apprentices and the absence of quantity limits imply that an *equilibrium* apprenticeship premium cannot exceed the hiring cost $\overline{\phi}$.

If $c < \overline{\phi}$, $S_t^{NG}(c) = \{0\}$: no entrepreneur would be available as employer of apprentices. In such a situation, the market will be in equilibrium if and only if $D_t\left(b^{NG}(c), c\right) = 0$, in which case no apprenticeship positions are exchanged. If the demand for apprenticeship positions at premium level $\overline{\phi}$ is positive, then $\overline{\phi}$ is an equilibrium apprenticeship premium and the associated demand $D_t(0, c) := \gamma_t\left(\mathcal{A}\left(0, \overline{\phi}\right)\right)$ will determine the amount of apprenticeship positions exchanged on the market in period t. As $b^{NG}\left(\overline{\phi}\right) = 0$, there would be no apprenticeship benefits in equilibrium.

Whenever $c \leq \overline{\phi}$, $b^{NG}(c) = 0$. Then

$$D_t(0, c) = \gamma_t(\mathcal{A}(0, c)) = \gamma_t\left(\left(\underline{p}(0, c), \overline{p}(0, c)\right)\right)$$

In absence of apprenticeship benefits, though, it is immediate to see from (1.28) and (1.37) that $\overline{p}(0, c)$ is strictly decreasing in c and that, conversely, $\underline{p}(0, c)$ is strictly increasing in c. Then demand must be at least weakly decreasing in c. It follows that either there exists an equilibrium apprenticeship premium $c^* < \overline{\phi}$ at which the apprenticeship market shuts down (and $D_t\left(0, \overline{\phi}\right) = 0$) or $D_t\left(0, \overline{\phi}\right) > 0$ and then $c^* = \overline{\phi}$ is the unique equilibrium apprenticeship premium.

The dynamics of the apprenticeship market under the no-guild regime is also pretty simply described. Because there can be no benefits in equilibrium, the occupational threshold is

constant through time and its value is, for all $t \in \mathbb{N}_0$,

$$\hat{p}_t = \hat{p}(0) = \frac{w - \mu_l \pi}{\Delta \pi} \tag{1.53}$$

Because agents in cohort C_0 make an unrestricted occupational choice and the occupational thresholds does not vary in time, no agent in *any* cohort will ever make a restricted occupational choice. Similarly, the two educational thresholds will be time-invariant, because benefits remain fixed at zero and, as long as the apprenticeship market exists, the apprenticeship premium remains fixed at $\overline{\phi}$. Thus, for all $t \in \mathbb{N}_0$,

$$\overline{p}_t = \overline{p}\left(0, \overline{\phi}\right) = \frac{-\overline{\phi} + \beta\rho\mu_h\left[w - \mu_\ell \pi\right]}{\beta\rho\Delta w} \tag{1.54}$$

$$\underline{p}_t = \underline{p}\left(0, \overline{\phi}\right) = \frac{\overline{\phi} + \beta\rho\mu_\ell\left[w - \mu_\ell \pi\right]}{\beta\rho\Delta\left(\pi - w\right)} \tag{1.55}$$

The apprenticeship pool will then be constant and, for all $t \in \mathbb{N}_0$, it coincides with

$$\mathcal{A} = \left(\underline{p}\left(0, \overline{\phi}\right), \overline{p}\left(0, \overline{\phi}\right)\right) \tag{1.56}$$

The equilibrium demand for apprenticeship positions in period $t \in \mathbb{N}_0$ is $D_t\left(0, \overline{\phi}\right) = \gamma_t\left(\mathcal{A}\right)$. The initial belief assignment p_0 determines $D_0\left(0, \overline{\phi}\right) = \gamma_0\left(\mathcal{A}\right)$. We make the following parametric assumption:

[A2] $\overline{\phi} < \frac{\beta\rho[\mu_h \pi - w][w - \mu_\ell \pi]}{\pi}$.

Assumption [A2] requires that the cost of hiring an apprentice in the no-guild regime, which coincides with the equilibrium apprenticeship premium as long as the apprenticeship market is active, be smaller than the level of apprenticeship premium beyond which the apprenticeship pool is empty, as it can be seen from (1.30) and (1.39). Under assumption [A2], $\mathcal{A}$ is non-empty. Therefore, by our assumption about the initial belief assignment, $D_0\left(0, \overline{\phi}\right) > 0$, and $c_0^* = \overline{\phi}$.

The key aspect to understand the dynamics of the demand for apprenticeship positions under the no-guild regime, and thus the dynamics of the entire market, is to notice that dynasties can exit but not enter the apprenticeship pool when the latter is constant. Moreover, a positive fraction of those dynasties who are in the apprenticeship pool in period t will not be there any longer in period $t+1$, because of the effects of learning. As we show in the Appendix, the apprenticeship pool can be partitioned into three subsets (essentially a lower, a middle and an upper portion) such that agents in the middle portion will exit the apprenticeship pool whenever they experience an effective apprenticeship;

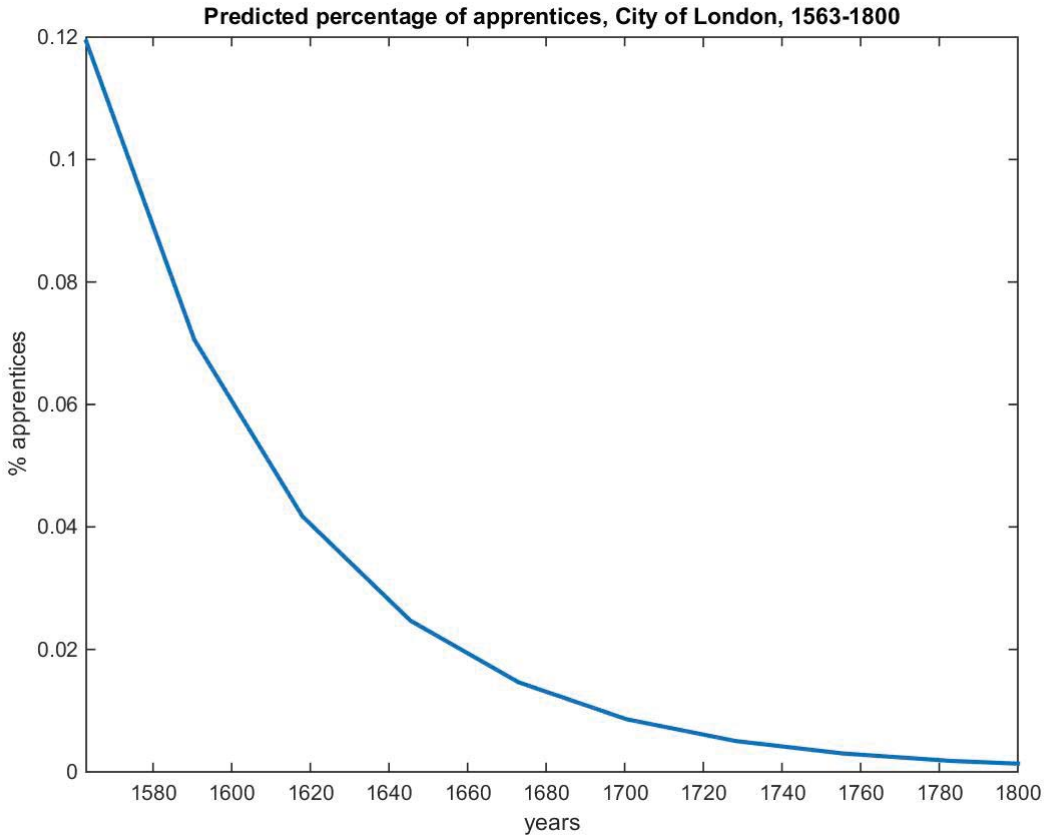


Figure 1.1: Evolution of equilibrium percentage of apprentices over total population (average of 1000 simulations) under a no-guild regime.

agents in the lower portion will exit the apprenticeship pool whenever they happen to be unsuccessful; agents in the upper portion will exit the apprenticeship pool whenever they happen to be successful. Because the elements of such a partitions are sub-intervals of $\mathcal{A}$, our assumption about the initial distribution of beliefs guarantees that there is a positive size of agents in each of the categories listed above. The total amount of dynasties in the apprenticeship pool is positive in each period, but strictly decreases, because the departure of agents by the above mechanism cannot be compensated by any form of entry from outside the apprenticeship pool: by construction, non-apprenticed agents cannot learn.

The consequences are immediate. Since in each period $\gamma_t(\mathcal{A}) > 0$, the apprenticeship market will never disappear: for all $t \in \mathbb{N}_0$, $c_t^* = \overline{\phi}$. But, as in each period a fraction of those in the apprenticeship pool have descendants whose belief are not in the appren-

ticeship pool, the aggregate demand for apprenticeship position will decrease from one generation to the next. The dynamics of learning then produces a steady contraction in the dimension of the apprenticeship market.

The graphical depiction in Figure 1.1 is obtained by repeated simulation of the evolution of the apprenticeship market under the no-guild regime. Details of the parametrization are given in the Appendix. As we showed in the present Section, the equilibrium percentage of apprenticeship in the population decreases at each step (10 generations have been used to produce Figure 1.1).

1.5.2 The apprenticeship market under the guild regime

Under the guild regime, a maximal individual intake of apprentices $m_{\max} \in \mathbb{N}$ is fixed and enforced by the authority. At the same time, the individual cost of hiring an apprentice is set to $\underline{\phi} \in \mathbb{R}_{++}$, with $\underline{\phi} < \overline{\phi}$. Unlike the no-guild case, because of the presence of a quantitative limit in the form of a maximal intake, the equilibrium apprenticeship premium under the guild regime might be larger than $\underline{\phi}$ and, therefore, apprenticeship benefits might be positive.

As seen in (1.51), each entrepreneur can supply the market with at most $m_{\max}$ apprenticeship positions, and will do so whenever $c > \underline{\phi}$. If $c < \underline{\phi}$, no entrepreneur will offer any apprenticeship position. Finally, when $c = \underline{\phi}$, each entrepreneur will be indifferent about how many apprenticeship positions to offer.

The function $\hat{p}$ introduced in (1.22) is strictly decreasing on $\mathbb{R}_{++}$. Using the definition of b^G, one finds that the function $c \mapsto \hat{p}\left(b^G(c)\right)$ is constant, and taking on the value $\hat{p}(0)$, when $c \leq \underline{\phi}$; while it is strictly decreasing when $c > \underline{\phi}$. It follows that, for all $t \in \mathbb{N}_0$, $e_t^G(c)$ is constant when $c \leq \underline{\phi}$, where it equals

$$e_t^G(0) = \gamma_t(\mathcal{E}(0))$$

while it is increasing when $c > \underline{\phi}$. On this domain, the supply correspondence (1.51) is, in fact, a function and, by our considerations, an increasing one. In particular, because of the initial belief assignment, $S_0^G(0)$ is *strictly* increasing on $\left(\underline{\phi}, +\infty\right)$.

Consider now the guild-regime demand function introduced in (1.52). For all $t \in \mathbb{N}_0$, if $c \leq \underline{\phi}$,

$$D_t^G(c) = \gamma_t(\mathcal{A}_t(0,c)) = \gamma_t\left(\left(\underline{p}(0,c), \overline{p}(0,c)\right)\right)$$

Notice from (1.37) that $\overline{p}(0,c)$ is strictly decreasing in c; similarly, from (1.28), we have that $\underline{p}(0,c)$ is strictly increasing in c. As c increases in $\left(0, \underline{\phi}\right)$, the interval $\mathcal{A}_t(0,c)$ shrinks and $\gamma_t\left(\mathcal{A}_t^G(c)\right)$ decreases. For the usual reason, D_0^G will be strictly decreasing on $\left(0, \underline{\phi}\right)$.

When $c > \underline{\phi}$, apprenticeship benefits under the guild regime are positive and the apprenticeship pool is determined by the bounds

$$\overline{p}\left(b^G(c), c\right) = \frac{-c + \beta\rho\mu_h\left[w - \left(\mu_\ell\pi + b^G(c)\right)\right]}{\beta\rho\Delta\left(w - b^G(c)\right)} \tag{1.57}$$

and

$$\underline{p}\left(b^G(c), c\right) = \frac{c + \beta\rho\mu_\ell\left[w - \left(\mu_\ell\pi + b^G(c)\right)\right]}{\beta\rho\Delta\left(\pi + b^G(c) - w\right)} \tag{1.58}$$

Both functions $c \mapsto \overline{p}\left(b^G(c), c\right)$ and $c \mapsto \underline{p}\left(b^G(c), c\right)$ are continuous in c on $\left(\underline{\phi}, +\infty\right)$ and they are also differentiable. In the Appendix we show that, for all $c \in \left(\underline{\phi}, +\infty\right)$, $\frac{d}{dc}\overline{p}\left(b^G(c), c\right) < 0$. Therefore $\overline{p}\left(b^G(c), c\right)$ is strictly decreasing on $\left(\underline{\phi}, +\infty\right)$. We also show that, for all $c \in \left(\underline{\phi}, +\infty\right)$,

$$\frac{d}{dc}\underline{p}\left(b^G(c), c\right) > 0 \iff m_{\max} \leq \frac{\pi - w}{\beta\rho\mu_\ell\mu_h\pi + \underline{\phi}} \tag{1.59}$$

We impose this condition as assumption

[A3] $m_{\max} \leq \frac{\pi - w}{\beta\rho\mu_\ell\mu_h\pi + \underline{\phi}}$

Under [A3], the demand function D_t^G is decreasing on $\left(\underline{\phi}, +\infty\right)$ as well and, therefore, decreasing on all $\mathbb{R}_+$. In particular, because of the initial assignment of beliefs, D_0^G is *strictly* decreasing on $\mathbb{R}_+$.

So far, we have shown that D_0^G is strictly decreasing on $\mathbb{R}_+$ and that S_0^G is a strictly increasing function on $\left(\underline{\phi}, +\infty\right)$. Under [A3], there exists a unique value $c_0^* \in \mathbb{R}_+$ such that the apprenticeship market is in equilibrium, i.e. such that $D_0^G(c_0^*) \in S_0^G(c_0^*)$. In particular,

- if $D_0^G\left(\underline{\phi}\right) \in \left[0, m_{\max}e_0^G\left(\underline{\phi}\right)\right]$, then $c_0^* = \underline{\phi}$ and there are no apprenticeship benefits in equilibrium;
- if $D_{t0}^G\left(\underline{\phi}\right) > m_{\max}e_0^G\left(\underline{\phi}\right)$, then $c_0^* > \underline{\phi}$ and equilibrium apprenticeship benefits are given by $b_0^* = b_0^G(c_0^*) = \left(c_0^* - \underline{\phi}\right)m_{\max}$.

The uniqueness of the equilibrium makes the aggregate belief transition (i.e. the measure change from γ_0 to γ_1) univocally defined. As before, for all $Q \in \mathcal{B}_{(0,1)}$, at least a fraction $(1-\rho)$ of agents in cohort C_0 will have descendants retaining the parental belief. This means that, for all $Q \in \mathcal{B}_{(0,1)}$, $\gamma_1(Q) \geq (1-\rho)\gamma_0(Q)$. By our assumption on the initial

belief assignment, this implies that, for all $Q \in \mathcal{B}_{(0,1)}$, $\gamma_1(Q) > 0$. But then D_1^G is also strictly decreasing on $\mathbb{R}_+$ and S_1^G is also strictly increasing on $\left(\underline{\phi}, +\infty\right)$. Thus the same equilibrium reasoning applies to period 1.

This line of reasoning can be replicated for any two consecutive periods t and $t+1$, showing by induction that a unique equilibrium is determined on the apprenticeship market for all $t \in \mathbb{N}_0$. That is, for all $t \in \mathbb{N}_0$ there exists $c_t^* \in \mathbb{R}_+$ such that $D_t^G(c_t^*) \in S_t^G(c_t^*)$ and the characterization of this equilibrium is identical to that of period 0 presented above.

It is also easy to replicate the reasoning proposed for the no-guild regime into showing that, for all $c \in \mathbb{R}_+$, $D_{t+1}^G(c) < D_t^G(c)$. Because of the effects of learning, in each period a positive measure of agents leave the *current* apprenticeship pool. Unlike under the no-guild regime, the new apprenticeship pool will differ from the previous one because a new equilibrium apprenticeship premium will be determined by the meet of demand and supply.

Assumption [A3] also implies that no possibility arises of there being constrained workers in the guild regime. Recall from Section 1.3 that a necessary and sufficient condition for the existence of constrained workers in period $t \in \mathbb{N}_0$ is that $\hat{p}_t < \underline{p}_{t-1}$. As $\underline{p}_{t-1} \leq \hat{p}_{t-1}$, this can only happen if $\hat{p}_t < \hat{p}_{t-1}$ which in turn requires that $c_t^* > c_{t-1}^*$. But under [A3] the last inequality also implies $\underline{p}_{t-1} < \underline{p}_t$. Hence we would have $\hat{p}_t < \underline{p}_t$, which only happens if c_t^* is above the limit in (1.30). But this can never happen in our equilibrium, because a large apprenticeship premium could only be justified by a large demand, which would result in a contradiction since no demand for apprenticeship positions can be expressed by agents for such a large premium.

1.5.3 Regime change

In this model the guild is an organization promoted by the community of entrepreneurs, which is sustained whenever it can guarantee positive apprenticeship benefits. The guild's action only concerns the apprenticeship market: the entrepreneurial profit π is not affected by the existence of the guild, nor does the latter modifies the individual probabilities of success which, as we have seen, only depend on dynastic traits. Equilibrium apprenticeship benefits can be positive only if, under the guild regime, the aggregate demand for apprenticeship positions is larger than the maximum supply of apprenticeship positions at $c = \underline{\phi}$. This then becomes a necessary and sufficient condition for the guild being supported.

Suppose this is the case in period 0. Because of the learning process, more and more dynasties will have their beliefs move towards the extrema 0 and 1 because of apprenticing.

Even though our model does not guarantee a strictly decreasing sequence of equilibrium apprenticeship premiums under the guild regime, this movement of masses of agents out of the subsequent apprenticeship pools will cause a decrease in equilibrium premiums through generations in the long run.

Decreasing equilibrium apprenticeship premiums will lead to decreasing equilibrium apprenticeship benefits. This will cause the occupational threshold $\hat{p}_t$ to increase in time, but the mass of entrepreneurs (i.e. the measure of those dynasties whose current member holds a belief $p > \hat{p}_t$) cannot decrease below the mass of those who initially held a belief above the largest possible upper bound of any apprenticeship pool in the guild regime, namely $\overline{p}\left(0, \overline{\phi}\right)$. Consequently, there must exist a time $T \in \mathbb{N}$ such that T is the smallest natural number at which $D_T^G\left(\underline{\phi}\right) < m_{\max} e_T^G\left(\underline{\phi}\right)$.

If the guild regime persisted, the equilibrium apprentice premium at time T would be $c_T^* = \underline{\phi}$. But at this point, the fact that $b_T^G\left(c_T^*\right) = 0$ induces a regime change by which the guild ceases to exist, paving the way for a no-guild regime. At this point two possibilities arise. If $D_T^{NG}\left(\overline{\phi}\right) = 0$, the apprenticeship market would shut down as soon as the regime change took place, because at the new level of individual hiring costs $\overline{\phi}$ no agent would be induced to educate her descendant, even though all entrepreneurs would be hiring any amount of apprentices. If, on the other hand, $D_T^{NG}\left(\overline{\phi}\right) > 0$, then the usual no-guild equilibrium, with $c_T^* = \overline{\phi}$ and a positive exchange of apprenticeship position would be established.

As demand keeps decreasing under the no-guild regime, a reversion to the old-guild regime would be impossible. Thus the no-guild regime acts here as an "absorbing state" for market regimes in this economy. Once the no-guild regime is established, no guild regime can ever take place again. In this case, the equilibrium apprenticeship premium would suddenly rise from the last guild-regime value c_{T-1}^* (presumably slightly larger than $\underline{\phi}$) to the no-guild regime value $\overline{\phi}$. If the difference between $\overline{\phi}$ and $\underline{\phi}$ were large enough, an observation of the evolution of apprenticeship premium should display an upward jump corresponding to the collapse of the guild regime.

Finally, notice that $\hat{p}_T$ would be larger than $\hat{p}_{T-1}$ because apprenticeship benefits in a no-guild equilibrium with $c_T^* = \overline{\phi}$ are zero while, by construction of the mechanism of guild sustainment, such benefits must have been positive in period $\hat{p}_{T-1}$. This also excludes the possibility that constrained workers appear in the first period subsequent to the establishment of the no-guild regime.

Figure 1.2 illustrates the phenomena described in this Subsection. To help the reader, this Figure makes use of a sort of "calibration" (described in Appendix B) that is able to set the regime shift at around the time we believe it actually happened (in the first quarter

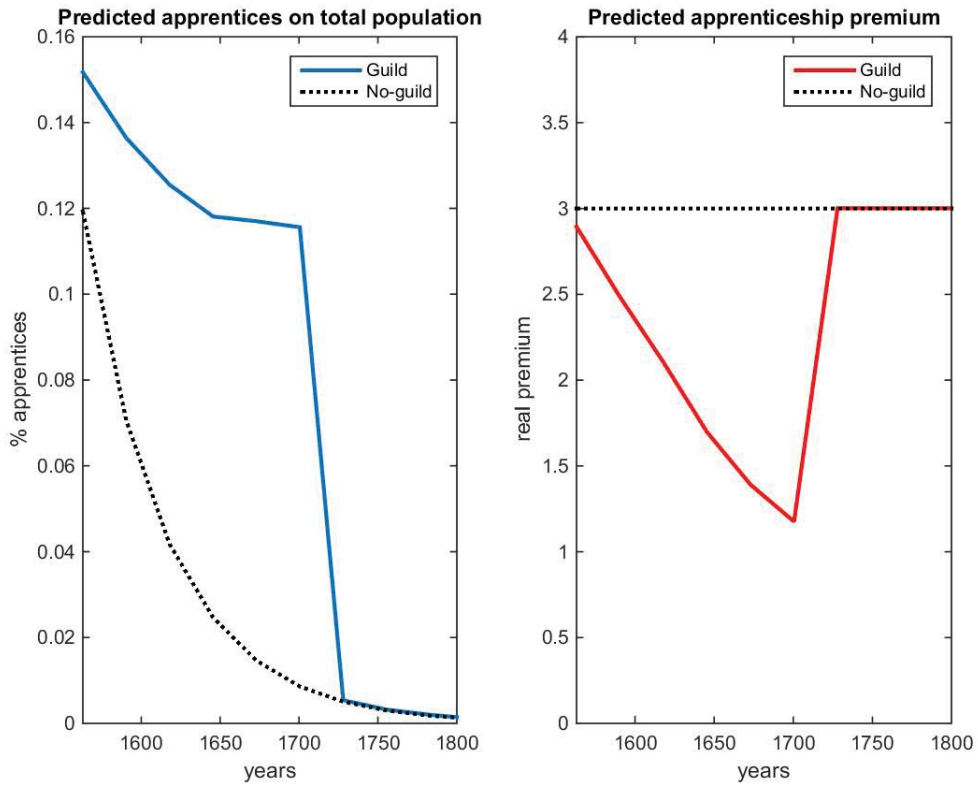


Figure 1.2: Predicted evolution of equilibrium percentage of apprenticeship over whole population and real apprenticeship premium (average of 1000 simulations).

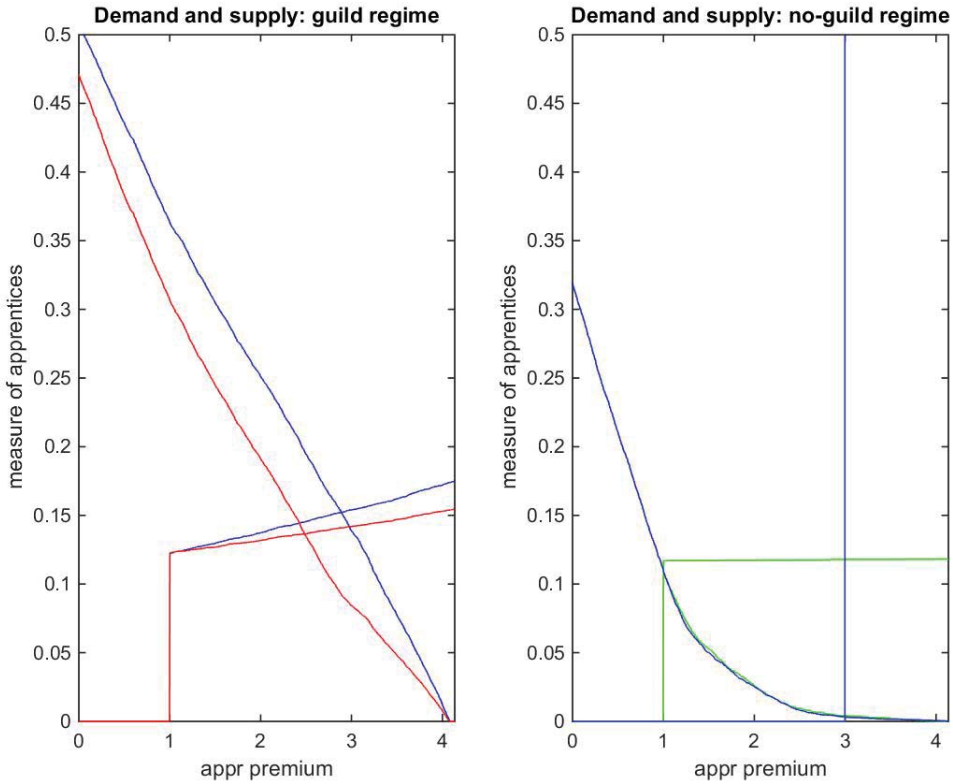


Figure 1.3: Left panel: demand and supply curves under the guild regime (blue: period 1; red: period 2). Right panel: theoretical demand and supply curves under the guild regime in period 8 (green) and actual demand and supply in period 8, under the no-guild regime (blue)

of the 18th century). Such calibration relies on arbitrary assumptions, e.g. about the initial distribution of beliefs, which cannot, of course, be justified by means of observable quantities.

The time span under observation, constituted of ten 25-years periods, can be divided into two phases. In a first phase the market for apprenticeship positions works under the guild-regime. The dynamics of individual learning yield a decrease in aggregate demand. As we have seen in the previous Subsection, the effect on supply is ambiguous. The left panel in Figure 1.3 represents two pairs of demand and supply schedules for the apprenticeship market under the guild regime. In particular, the blue lines are the demand and supply curves for apprenticeship positions in period 1 of the simulation and the red lines are the demand and supply curves for apprenticeship positions in the subsequent period 2. Their intersections determine the equilibrium apprenticeship premium in those periods. As one

can see, for any level of prices, the period-2 demand for apprenticeship positions is smaller than the period-1 demand. Supply of apprenticeship positions in period 2 is larger than supply in period 1 for values of the premium just above $\underline{\phi} = 1$ and then becomes smaller in correspondence to larger values of the premium.

In a second phase, beginning around 1720 in our simulation, the guild cannot guarantee positive profits from apprenticeship. As a consequence, the market shifts towards a no-guild regime. This situation is depicted in Figure 1.3. The green lines are the *theoretical* demand and supply curve that would characterize the apprenticeship market in period 8 of our simulation (1755-1783). One notices that demand at $\underline{\phi} = 1$ is smaller than supply; consequently, the guild-regime equilibrium apprenticeship premium would be $c^* = 1$ and a guild would not be able to guarantee positive apprenticeship benefits to its associates. Therefore, the market shifts towards a no-guild regime and $c^* = \overline{\phi} = 3$. The actual demand and supply determining the market in period 8 are shown in blue in the same panel. Supply is indeterminate at $c = \overline{\phi} = 3$; as one can see, the no-guild-regime demand curve coincides with the guild-regime demand curve for $c \leq \underline{\phi}$.

The regime shift also characterizes the evolution of the occupational and educational thresholds, represented in Figure 1.4. The decreasing apprenticeship premium under the guild regime produces an increase in the occupational threshold (as apprenticeship benefits are reduced, entrepreneurship becomes less and less convenient) and a progressively larger apprenticeship pool (because, by our assumption A3, the upper educational threshold is decreasing in the premium and the lower educational threshold is increasing in the premium). As the regime shift takes place, educational and occupational thresholds stabilize at their no-guild levels and remain constant thereafter.

1.5.4 Historical evidence

The evolution of the apprenticeship market (in terms of its dimension with respect to the total population) and of the apprenticeship premium suggested by our model are compatible with the observations available for the period under study. Figure 1.5 compares the model's predictions about apprentices' share on total population (top panel) and apprenticeship premium (bottom panel), both represented in red, with the series obtained from recorded evidence, indicated by the blue lines. Because of the numerical assumptions used in generating the model's predictions, we must once again underline that these comparisons ought to be read for their qualitative content, not in a quantitative spirit. Nonetheless, the major discrepancies between prediction and reality in the top panel can be easily explained.

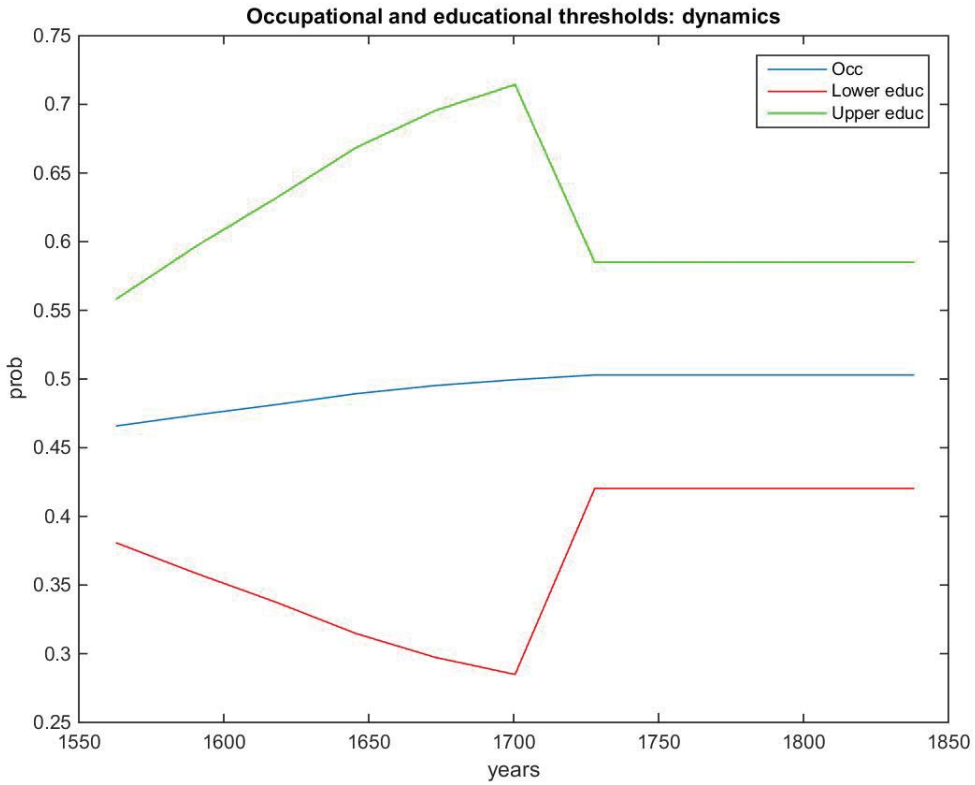


Figure 1.4: Dynamics of occupational and educational thresholds.

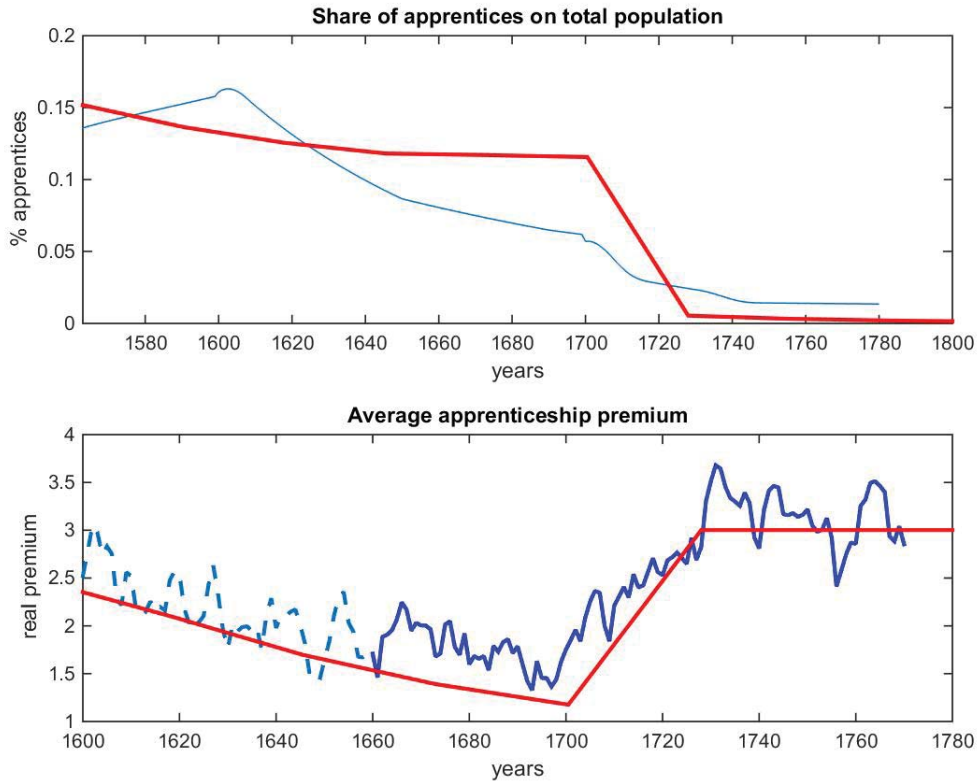


Figure 1.5: Top panel: evolution of the percentage of apprentices over total population (City of London, blue line) and same percentage as predicted by the model (from Fig. 1.2). Bottom panel: evolution of the average real apprenticeship premium (City of London, blue line; the dashed portion indicates values obtained by assigning constant nominal premiums) and apprenticeship premium predicted by the model (from Fig. 1.2).

First of all, the model features a sudden regime shift between 1700 and 1720. While data confirm that this was a period of intense decline in the apprenticeship market's dimensions, we see that such decline had been characterizing the whole previous century. While the model, because of its simplifications, is drastic in predicting the regime shift, reality was much more gradual. Guilds did not suddenly lost power, but their grip faded away slowly. According to Kellett (1958) and Epstein and Prak (2008) the decline of guilds must have taking place between 1650 and 1720, with the last 10-15 years being characterized by a fastening of its pace.

Secondly, the amount of apprentices under the no-guild regime is smaller in our model than in observations. This is because the model purposefully ignores a physiological mass of (real) agents that would enter apprenticeship outside of the logic that motivates our agents' choices. In particular, our model ignores population growth and assigns one descendant to each agent. By so doing, we ignore the (real) parent's problem of allocating children across professions. The average number of children per urban family was around 2,5-3 in the period of our observation (Finlay, 1981). Even when apprenticeship had lost most of its value with the advent of the no-guild regime, some families may have sent *some* of their children to be apprenticed for lack of better opportunities; or just in order to reproduce a family's activity in time. Finally, it is also possible that some of the recorded apprentices after 1730 belong to the array of "poor apprentices", who began in some instances to be registered as premium-payers around those years. All this may contribute to the discrepancy we observe in the top panel of Figure 1.5 for the no-guild years.

1.6 Conclusions

The objective of this paper was to suggest that professional training can be a suitable means for agents to discover unobservable characteristics of their own, which can help them make better-informed decisions once the training is over. In particular, we wished to interpret in this direction the important modern-age institution of apprenticeship. In order to do so, we built a model of occupational and educational choice characterized by the intergenerational transmission of a fixed, yet unobservable trait, which determines individual chances of success in an entrepreneurial activity. The main idea is that entrepreneurial success depends on a vast array of individual characteristics, many of which might be unknown to those who own them. Professional education can help agents uncover such characteristics and, by letting them make better-informed choices, contribute to social efficiency.

The model has several interesting implications. For instance, it predicts that, as long as the educational institution (in our case, apprenticeship) persists, the social background of participants (in terms of parental occupations) will be diverse. Moreover, the aggregate implications of our model are in line with the available evidence on modern-age apprenticeship in the City of London. In particular, the model provides an explanation for the endogenous decrease in the size of the apprenticeship market and for the dynamics of average apprenticeship premiums (representing the main cost of education) appearing in available data.

Because our model does not delve with the microeconomics of the apprenticeship contract, this paper does not contest the plausibility of the hold-up theory about the seemingly excessive length of the training period. On the other hand, though, it may be read as a suggestion that the necessities of learning without being effectively taught might have contributed to such feature. The aggregate role of apprenticeship as a selector of workers' quality, which characterizes the quality-signaling interpretation of this institution, is also reflected by our implication that, through learning, those who are most likely to succeed as entrepreneurs will, in the long run, choose this occupation. In the meantime, we do not deny the distortive intentions of modern-age guilds, although dynamic aspects of guild intervention are necessarily neglected in our framework.

1.7 Appendices

1.7.1 Appendix A: Details

Law of Large Numbers on subsets of agents

In Section 1.3, we chose to represent the population of our agents by a countably infinite set $\mathcal{I}$ endowed with a probability charge λ. The reason for this choice is that we want to make use of a Law of Large Numbers that applies to subsets of agents. Suppose that, for some $t \in \mathbb{N}_0$, $(Z_i^t)_{i\in\mathcal{I}}$ is an *independent* stochastic process such that, for all $i \in \mathcal{I}$, Z_i^t has a well-defined expected value μ_i. Suppose also that the function $i \mapsto \mu_i$ has a well-behaved distribution function and asymptotic frequencies (see Al-Najjar, 2004, for the definitions). This is, for instance, the case when such function can take on at most two distinct values, which will be the case for this model. Let now A be an interval in $[0, 1]$. Then it holds (almost everywhere on the probability space on which every r.v. Z_i^t is defined) that

$$\int_A Z_i^t d\lambda = \int_A \mu_i d\lambda \tag{1.60}$$

In particular, suppose that, for all $i \in \mathcal{I}$, $Z_i^t \sim Ber(\mu_i)$ where $\mu_i = \mu_\ell$ if $\theta^i = \ell$ and $\mu_i = \mu_h$ if $\theta^i = h$. Suppose also that the belief assignment p_t at time t is given. We can re-order agents through a permutation (which we abstain from indicating explicitly for simplicity) so that $i < i'$ if and only if $p_t^i < p_t^{i'}$. Then, for an *interval* $Q \subseteq (0,1)$, the set of agents with belief in Q in period t, that is $p_t^{-1}(Q)$, is an interval in $[0,1]$ under said permutation. Result (1.60) allows us to say that the mass of agents with beliefs in Q for which $Z^t = 1$ is equal to the average of their individual expected values μ_i, and therefore that

$$\int_{p_t^{-1}(Q)} Z_i^t d\lambda = \int_{p_t^{-1}(Q)} \mu_i d\lambda = \mu_h \varphi_t(h, Q) + \mu_\ell \varphi_t(\ell, Q) \tag{1.61}$$

Notice (and refer to Al-Najjar, 2004, as always for details) that these results crucially depend on $\mathcal{I}$ being countable and on λ being a probability charge on $\left(\mathcal{I}, 2^{\mathcal{I}}\right)$.

Working parents' educational choice

In this Subsection we derive conditions (1.27) and (1.30). A working parent will apprentice her son if and only if $U_{i_t}(W, A) > U_{i_t}(W, NA)$. By (1.26), this holds if and only if

$$\beta\rho\mu\left(p^{i_t}\right)\left[\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t - w\right] > c_t \tag{1.62}$$

Notice that

$$\begin{aligned}
\mu\left(p^{i_t}\right)\mu\left(q^+\left(p^{i_t}\right)\right) &= \mu\left(p^{i_t}\right)\left(\frac{\mu_h p^{i_t}}{\mu\left(p^{i_t}\right)}\Delta + \mu_\ell\right) \\
&= \mu_h p^{i_t}\Delta + \mu_\ell\left(p^{i_t}\Delta + \mu_\ell\right) \\
&= p^{i_t}\Delta + \mu_\ell^2
\end{aligned} \tag{1.63}$$

Then (1.62) becomes

$$\begin{aligned}
\beta\rho\left[\left(p^{i_t}\Delta + \mu_\ell^2\right)\pi + \left(p^{i_t}\Delta + \mu_\ell\right)(b_t - w)\right] &> c_t \\
p^{i_t}\Delta\left[\beta\rho(\pi + b_t - w)\right] + \beta\rho\mu_\ell\left[\mu_\ell\pi + b_t - w\right] &> c_t \\
\frac{c_t + \beta\rho\mu_\ell\left[w - (\mu_\ell\pi + b_t)\right]}{\beta\rho\Delta(\pi + b_t - w)} &< p^{i_t}
\end{aligned}$$

which is (1.27). The absence of apprenticeship by working parents occurs when the lower educational threshold identified by (1.27) is at least as large as the occupational threshold.

That is whenever

$$\begin{aligned}
\underline{p}\left(b_t, c_t\right) &\geq \hat{p}\left(b_t\right) \\
\frac{c_t + \beta\rho\mu_\ell\left[w - \left(\mu_\ell\pi + b_t\right)\right]}{\Delta\beta\rho\left(\pi + b_t - w\right)} &\geq \frac{w - \left(\mu_\ell\pi + b_t\right)}{\Delta\pi} \\
c_t\pi + \beta\rho\mu_\ell\pi\left[w - \left(\mu_\ell\pi + b_t\right)\right] &\geq \beta\rho\left(\pi + b_t - w\right)\left[w - \left(\mu_\ell\pi + b_t\right)\right] \\
c_t\pi &\geq \beta\rho\left[w - \left(\mu_\ell\pi + b_t\right)\right]\left[\left(\mu_h\pi + b_t\right) - w\right] \\
c_t &\geq \frac{\beta\rho\left[w - \left(\mu_\ell\pi + b_t\right)\right]\left[\left(\mu_h\pi + b_t\right) - w\right]}{\pi}
\end{aligned}$$

which yields condition (1.30).

Entrepreneur parents' educational choice

In this Subsection we derive conditions (1.36) and (1.39). In order to derive (1.35) just proceed as follows from (1.34):

$$\begin{aligned}
U_{i_t}\left(E, A\right) &= \mu\left(p^{i_t}\right)\pi + b_t - c_t + \\
&+ \beta\left\{\left(1-\rho\right)\left(\mu\left(p^{i_t}\right)\pi + b_t\right) + \right. \\
&+ \left.\rho\left[\mu\left(p^{i_t}\right)\left(\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t\right) + \left(1 - \mu\left(p^{i_t}\right)\right)w\right]\right\} \\
&= \mu\left(p^{i_t}\right)\pi + b_t - c_t + \\
&+ \beta\left\{\left(1-\rho\right)\left(\mu\left(p^{i_t}\right)\pi + b_t\right) + \right. \\
&+ \left.\rho\left[\mu\left(p^{i_t}\right)\left(\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t\right) \pm \left(\mu\left(p^{i_t}\right)\pi + b_t\right) + \left(1 - \mu\left(p^{i_t}\right)\right)w\right]\right\} \\
&= \left(1+\beta\right)\left(\mu\left(p^{i_t}\right)\pi + b_t\right) - c_t + \\
&+ \beta\rho\left[\mu\left(p^{i_t}\right)\left(\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t\right) - \left(\mu\left(p^{i_t}\right)\pi + b_t\right) + \left(1 - \mu\left(p^{i_t}\right)\right)w\right] \\
&= \left(1+\beta\right)\left(\mu\left(p^{i_t}\right)\pi + b_t\right) - c_t + \\
&+ \beta\rho\left[\mu\left(p^{i_t}\right)\left(\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t - \pi - w\right) + w - b_t\right] \\
&= U_{i_t}\left(E, NA\right) - c_t + \\
&+ \beta\rho\left[\mu\left(p^{i_t}\right)\left(\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t - \pi - w\right) + w - b_t\right]
\end{aligned}$$

Then an entrepreneur parent will apprentice her son if and only if

$$\begin{aligned}
\beta\rho\left[\mu\left(p^{i_t}\right)\left(\mu\left(q^+\left(p^{i_t}\right)\right)\pi + b_t - \pi - w\right) + w - b_t\right] &> c_t \\
\beta\rho\left[\left(p^{i_t}\Delta + \mu_\ell^2\right)\pi + \left(p^{i_t}\Delta + \mu_\ell\right)\left(b_t - \pi - w\right) + w - b_t\right] &> c_t \\
\beta\rho\left[p^{i_t}\Delta\left(b_t - w\right) - \mu_\ell\mu_h\pi + \mu_h\left(w - b_t\right)\right] &> c_t \\
p^{i_t}\left[\beta\rho\Delta\left(w - b_t\right)\right] &< -c_t + \beta\rho\mu_h\left[w - \left(\mu_\ell\pi + b_t\right)\right] \\
\frac{-c_t + \beta\rho\mu_h\left[w - \left(\mu_\ell\pi + b_t\right)\right]}{\beta\rho\Delta\left(w - b_t\right)} &< p^{i_t}
\end{aligned}$$

which gives (1.36). The absence of apprenticeship by entrepreneur parents occurs when the upper educational threshold identified by (1.36) is no larger that the occupational threshold. That is, whenever

$$\begin{aligned}
\overline{p}\left(b_t, c_t\right) &\leq \hat{p}\left(b_t\right) \\
\frac{-c_t + \beta\rho\mu_h\left[w - \left(\mu_\ell\pi + b_t\right)\right]}{\beta\rho\Delta\left(w - b_t\right)} &\leq \frac{w - \left(\mu_\ell\pi + b_t\right)}{\Delta\pi} \\
-c_t\pi + \beta\rho\mu_h\pi\left[w - \left(\mu_\ell\pi + b_t\right)\right] &\leq \beta\rho\left(w - b_t\right)\left[w - \left(\mu_\ell\pi + b_t\right)\right] \\
c_t\pi &\geq \beta\rho\left[w - \left(\mu_\ell\pi + b_t\right)\right]\left[\left(\mu_h\pi + b_t\right) - w\right] \\
c_t &\geq \frac{\beta\rho\left[w - \left(\mu_\ell\pi + b_t\right)\right]\left[\left(\mu_h\pi + b_t\right) - w\right]}{\pi}
\end{aligned}$$

which yields condition (1.39).

Dynamics of the apprenticeship market under the no-guild regime

Consider the time-invariant apprenticeship pool $\mathcal{A} = \left(\underline{p}\left(0, \overline{\phi}\right), \overline{p}\left(0, \overline{\phi}\right)\right)$. By assumption [A2], $\mathcal{A} \neq \emptyset$. Recall that, for all $b \in \mathbb{R}_+$, $\underline{p}\left(b, 0\right) = q^-\left(\hat{p}\left(b\right)\right)$ and $\overline{p}\left(b, 0\right) = q^+\left(\hat{p}\left(b\right)\right)$. Moreover, recall that $\underline{p}\left(0, \cdot\right)$ is strictly increasing in c and $\overline{p}\left(0, \cdot\right)$ is strictly decreasing. Since $\overline{\phi} > 0$, then $\mathcal{A} \subsetneq \left(q^-\left(\hat{p}\left(0\right)\right), q^+\left(\hat{p}\left(0\right)\right)\right)$. In particular, by the properties of q^- and q^+, $\underline{p}\left(0, \overline{\phi}\right) > q^-\left(\hat{p}\left(0\right)\right)$ implies $q^+\left(\underline{p}\left(0, \overline{\phi}\right)\right) > \hat{p}\left(0\right)$ and, analogously, $\overline{p}\left(0, \overline{\phi}\right) < q^+\left(\hat{p}\left(0\right)\right)$ implies $q^-\left(\overline{p}\left(0, \overline{\phi}\right)\right) < \hat{p}\left(0\right)$. It follows that $\mathcal{A}$ can be partitioned as follows:

$$\mathcal{A} = \left(\underline{p}\left(0, \overline{\phi}\right), q^-\left(\overline{p}\left(0, \overline{\phi}\right)\right)\right) \sqcup \left[q^-\left(\overline{p}\left(0, \overline{\phi}\right)\right), q^+\left(\underline{p}\left(0, \overline{\phi}\right)\right)\right] \sqcup \left(q^+\left(\underline{p}\left(0, \overline{\phi}\right)\right), \overline{p}\left(0, \overline{\phi}\right)\right)$$

We call these three sets the (respectively, in the order they are written) lower, middle and upper portion of $\mathcal{A}$ and we denote them by $\mathcal{A}_L$, $\mathcal{A}_M$ and $\mathcal{A}_U$. Notice that $\mathcal{A}_M \neq \emptyset$ independently of the value of $\overline{\phi}$, while it is possible that $\mathcal{A}_L$ and/or $\mathcal{A}_U$ be empty.

$\mathcal{A}_L$ is the set of beliefs with respect to which updating downwards implies leaving the apprenticeship pool while updating upwards implies remaining in the apprenticeship pool. Conversely, agents with beliefs in $\mathcal{A}_U$ will leave the apprenticeship pool if they update upwards and will remain in the apprenticeship pool if they update downwards. Finally, agents with beliefs in $\mathcal{A}_M$ will leave the apprenticeship pool whenever they update their beliefs.

Now define, for each $t \in \mathbb{N}_0$ and each $Q \in \mathcal{B}_{(0,1)}$,

$$\overline{\mu}_t\left(Q\right) := \mu_h\varphi_t\left(h, Q\right) + \mu_l\varphi_t\left(\ell, Q\right) \tag{1.64}$$

Then $\overline{\mu}_t\left(Q\right)$ is the average probability of success of members of cohort C_t whose beliefs lie in Q. If Q is a sub-interval of $(0, 1)$, equation (1.61) guarantees that this is also the

theoretical size of the set of members of cohort C_{t+1} with inherited beliefs in Q that would *actually* be successful apprentices, if they were all effectively apprenticed. The implication is that, if $Q \subseteq \mathcal{A}$ (as it will be the case in our application), then the size of the actually successful apprentices among members of C_{t+1} with inherited beliefs in Q is $\rho\overline{\mu}_t(Q)$.

Consider now cohort C_t and the distribution of beliefs pertaining to it. The size of apprenticing parents in cohort C_t (i.e. the size of members of C_{t+1} who are apprenticed) is

$$\gamma_t(\mathcal{A}) = \gamma_t(\mathcal{A}_L) + \gamma_t(\mathcal{A}_M) + \gamma_t(\mathcal{A}_U)$$

Recall that, for all $j \in \{L, M, U\}$, if $\mathcal{A}_j \neq \emptyset$, then $\mathcal{A}_j$ is an interval and $\gamma_0(\mathcal{A}_j) > 0$. Consider first the agents whose beliefs at time 1 belongs to $\mathcal{A}_M$. Every agent in C_1 with inherited belief in this set at time 0 remains into the apprenticeship pool if and only if her apprenticeship is ineffective. No other agent with inherited beliefs in the period-0 apprenticeship pool can have produced a posterior belief within this set. Then $\gamma_1(\mathcal{A}_M) = (1-\rho)\gamma_0(\mathcal{A}_M)$. Then it is immediate to see that, for all $t \in \mathbb{N}$, $\gamma_t(\mathcal{A}_M) = (1-\rho)^t \gamma_0(\mathcal{A}_M)$ and, therefore, that $\gamma_t(\mathcal{A}_M) \to 0$ as $t \to +\infty$.

Agents whose belief at time 1 belongs to $\mathcal{A}_L \sqcup \mathcal{A}_U$ are either (1) descendants of parents whose belief at time 0 belonged to the same set and whose apprenticeship was not effective or (2) descendants of parents whose belief at time 0 belonged to $\mathcal{A}_L$ and who were successful apprentices or (3) descendants of those whose belief at time 0 belonged to $\mathcal{A}_U$ and who were failed apprentices. Then

$$\gamma_1(\mathcal{A}_L \sqcup \mathcal{A}_U) = (1-\rho)\gamma_0(\mathcal{A}_L \sqcup \mathcal{A}_U) + \rho\overline{\mu}_0(\mathcal{A}_L) + \rho\left[\gamma_0(\mathcal{A}_U) - \overline{\mu}_0(\mathcal{A}_U)\right]$$

Notice that, for all $t \in \mathbb{N}_0$ and for all sets $Q \in \mathcal{B}_{(0,1)}$ such that, for all $\tau \in \{\ell, h\}$, $\varphi_t(\tau, Q) > 0$, $\overline{\mu}_t(Q) \in (\mu_\ell\gamma_t(Q), \mu_h\gamma_t(Q))$ since $\mu_h > \mu_\ell$. Consequently, the measure of agents whose beliefs lie in Q and who are successful apprentices is smaller than $\rho\mu_h\gamma_t(Q)$ and larger than $\rho\mu_\ell\gamma_t(Q)$. Because $\mu_h = 1 - \mu_l$, the measure of agents whose beliefs lie in Q and who are failed apprentices obeys the same rule. Then

$$\begin{aligned}\gamma_1(\mathcal{A}_L \sqcup \mathcal{A}_U) &< (1-\rho)\gamma_0(\mathcal{A}_L \sqcup \mathcal{A}_U) + \rho\mu_h\gamma_0(\mathcal{A}_L \sqcup \mathcal{A}_U) \\ &= (1-\rho+\rho\mu_h)\gamma_0(\mathcal{A}_L \sqcup \mathcal{A}_U)\end{aligned}$$

But then, for all $t \in \mathbb{N}$,

$$\gamma_t(\mathcal{A}_L \sqcup \mathcal{A}_U) < (1-\rho+\rho\mu_h)^t \gamma_0(\mathcal{A}_L \sqcup \mathcal{A}_U)$$

implying, since $1-\rho+\rho\mu_h < 1$, that $\gamma_t(\mathcal{A}_L \sqcup \mathcal{A}_U) \to 0$ as $t \to +\infty$. We can then finally conclude that $\gamma_t(\mathcal{A}) \to 0$ as $t \to +\infty$.

Apprenticeship market under the guild regime

We show that the upper educational threshold is decreasing in the apprenticeship premium when the latter exceeds c. Then we derive the condition under which the lower educational thresholds is increasing in the apprenticeship premium on the same domain.

Differentiating $c \mapsto \overline{p}\left(b^G(c), c\right)$ with respect to c, one has

$$\frac{d}{dc}\overline{p}\left(b^G(c), c\right) = -\frac{m_{\max}\left(c + \beta\rho\mu_h\mu_\ell\pi\right) + w - b^G(c)}{\beta\rho\Delta\left(w - b^G(c)\right)^2}$$

Because $b^G(c) < w$ in any sensible situation allowed by this model, $\frac{d}{dc}\overline{p}\left(b^G(c), c\right) < 0$. We can similarly differentiate $c \mapsto \underline{p}\left(b^G(c), c\right)$ and find

$$\frac{d}{dc}\underline{p}\left(b^G(c), c\right) = \frac{\left(\pi + b^G(c) - w\right) - m_{\max}\left(\beta\rho\mu_\ell\mu_h\pi + c\right)}{\beta\rho\Delta\left(\pi + b^G(c) - w\right)^2}$$

At this point, we cannot assess the sign of this derivative. Notice though that, since $c > \underline{\phi}$ implies $m^G(c) = m_{\max}$ and hence $b^G(c) = \left(c - \underline{\phi}\right) m_{\max}$,

$$\left(\pi + b^G(c) - w\right) - m_{\max}\left(\beta\rho\mu_\ell\mu_h\pi + c\right) = (\pi - w) - m_{\max}\left(\beta\rho\mu_\ell\mu_h\pi + \underline{\phi}\right)$$

so that

$$\frac{d}{dc}\underline{p}\left(b^G(c), c\right) > 0 \iff m_{\max} \leq \frac{\pi - w}{\beta\rho\mu_\ell\mu_h\pi + \underline{\phi}}$$

1.7.2 Appendix B: Data and Simulations

Data

Data on the population of the City of London are obtained from the "Locating London's Past" project's population database (see www.locatinglondon.org) for the late 17th and 18th centuries and from Finlay (1981) for previous decades. Locating London's Past data allow us to isolate the population of the City, as numbers are disaggregated at the parish level. In order to harmonize this series with the numbers in Finlay (1981) we apply a proportionality constant to account for the fact that the latter uses evidence from the whole London area and not just from the City. The complete series is obtained by interpolating among pairs of consecutive observations and assuming a constant growth rate.

The estimated number of active apprentices is a re-elaboration of Schwarz' (1987) calculations. Using the author's data I computed, for each year, the estimated number

of apprentices based on the author's hypothesis that 10% of the apprenticing workforce would leave the training program for each of the 7 years. The data on the apprenticeship premiums are derived from the observations reported by Lane (1996). Also accounting for the facts that many of the high premiums there rely on Defoe's accounts, which are known to be heavily biased against the guild system, the average nominal premiums in the late 1600s appears to have ranged from around 5£for low-tier crafts to 60£for high-tier trades and liberal professions. We then construct an average indicator, distinguishing professions into two categories as above, using the by-profession data on premiums reported by Minns and Wallis (2013). We claim that the resulting average magnitude can be used for the years following the Great Fire. As for the preceding decades, we apply the same nominal values for lack of better observations, as a transitory solution. The reader should be warned that this operation stands in the realm of the tentative. The rest of the series is obtained from Minns and Wallis (2013).

Data about prices, wages and profits are obtained from the database associated with Allen (2001). In particular, we use his Consumer Price Index for London expressed in silver prices. All series used in this paper are expressed relative to this index. We then use the two series constructed by Allen for the daily wages of, respectively, agricultural laborers and urban building laborers, to obtain a unique series of average silver wages for the City of London and its rural surroundings. In order to do so, we averaged the two series using as weights the average allocation of workforce across primary and non-primary sector as illustrated by Shaw-Taylor and Wrigley (2008). The only profit series available in Allen (2001) is that for builder craftsmen. Fortunately builders were considered an average craft, not as humble as (say) blacksmiths, nor as noble as liberal arts practitioners. Hence we feel there is no dramatic limitation in using this series as an indicator of average silver profits.

Simulations

The simulations referred to in Section 1.5 are carried out under a certain number of parametric assumptions. Some of those are historically motivated; others have been arbitrary imposed, as a sort of calibration, mainly in order to let the simulated regime change coincide temporally with the historical one. For this reason, the role one should assign to such results is exclusively qualitative and no quantitative consideration should be based on them.

The fixed values of wages (w) and craft profits (π) for the simulations have been obtained from the data by Allen (2001). They coincide with the average yearly wages and profits

multiplied by the length that our model attributes to one's career, i.e. the length of a generation that, on the basis of the information provided by Finlay (1981) and Wrigley et al. (1989), has been set to 25 years. We set $\mu_h = 0.8$, $\beta = 0.9$, $\rho = 0.6$. In particular, notice that, if we believe in the theory proposed by this paper, the percentage of incomplete apprenticeships cannot be used as a proxy for apprenticeship *in*effectiveness: on the contrary, one can think that most of those who abandon apprenticeship do so because they have learned not to be excessively fit for entrepreneurship. We let $\underline{\phi} = 1$ and $\overline{\phi} = 3$. Costs sustained when hiring an apprenticeship have thus the same order of magnitude as yearly wages and yearly profits. Finally, we set $m_{\max} = 1$. This is due to the fact that, in the model, one period of apprenticeship lasts as a whole generation, i.e. 25 years, while in fact each apprenticeship contract had an average duration of 7 years. Thus each of our apprentices stands for around 3.5 historical apprentices. Since masters hired 3 to 4 apprentices at a time, as prescribed by the Statute, we set $m_{\max} = 1$ by a criterion of proportionality.

Chapter 2

Resistance to institutions and cultural distance: brigandage in post-unification Italy

Co-authored with Giampaolo Lecce and Laura Ogliari.

2.1 Introduction

The recent comeback of separatist and secessionist movements in Europe can be interpreted as a signal of the lack of cultural convergence between regions of the same country, as documented by Alesina et al. (2017). This interpretation suggests that the historical process of nation building, which led to the formation of contemporary European states by the end of the 19th century, has not always been successful. In many cases, nation building amounted to national rulers introducing uniform institutions across their territories and implementing policies intended to create a homogeneous legal environment and, possibly, a unified national identity.[1] Thus, incomplete nation building can often be attributed to flaws in the introduction of new institutions, such as the lack of acceptance of a newly imposed institutional framework.

In this paper, we investigate whether cultural differences play a relevant role in the short-run rejection of new institutions. This rejection may hinder state formation processes in the short run and can also have long-lasting effects on important aspects of nation building, such as the intensity of political participation. We address this issue by analyzing

[1] See Alesina and Reich (2015) for historical examples of nation building in Europe.

the consequences of the imposition of Piedmontese institutions on the territories in continental Southern Italy[2] during the Italian unification process (1860–1870). Three features of the Italian unification make it an ideal setting for our analysis. First, institutional rejection took on the very visible form of violent, guerrilla-like popular unrest known as *brigandage*. Second, historical migrations provide us with local variation in cultural similarity with Piedmont across southern municipalities. Finally, the incompleteness of Italy's nation building process is reflected by large and persistent differences in attitudes toward institutions, the intensity of which may be partially evaluated in light of this case of institutional rejection.

The contribution of our paper is threefold: first, we construct a new and unique dataset including information on recorded brigandage episodes at the municipal level and on pre- and post-unification socio-economic variables, thereby providing the first detailed and comprehensive quantification of the intensity of institutional rejection in the form of brigandage; second, we relate this intensity to cultural characteristics and find strong evidence that institutional rejection is closely linked to cultural distance from the environment in which the institutions originated; third, we highlight a persistent relationship between the intensity of the institutional rejection and the dynamics of political participation following the Italian unification.

We measure cultural distance from Piedmont by the geographical distance from communities descending from 'near-Piedmontese' settlers, whose persistent cultural similarity with their ancestors is well documented by historians.[3] Prolonged interactions with Piedmontese descendants, occasional intermarriages and protracted exposure to Piedmontese social norms will increase the cultural proximity of local communities to these near-Piedmontese settlements and, therefore, to Piedmont itself. Knowing the number of brigandage episodes in each municipality, we directly test whether cultural distance affects the strength with which communities reject transplanted institutions. Our results show that, on average, doubling the distance from the closest Piedmontese enclave is associated with a 17% increase in the expected incidence of brigand uprisings. We substantiate our claim of a social interaction channel by discussing anecdotal evidence of historical trade and intermarriages across neighboring communities and by showing that the effect of cultural proximity is non-linear and sensitive to the intensity of exposure to Piedmontese culture.

[2]From now on, we shall simply refer to this region as 'Southern Italy', but the limitation to the continental regions applies everywhere.

[3]We call 'near-Piedmontese' those communities established by migrants from the Provence-Savoy-Piedmont area. For further details, see Section 2.2.

Our results survive an extensive sensitivity analysis and are robust to the inclusion of controls for a wide range of alternative explanations (e.g., human and social capital, land use, financial development, etc.). A potential concern is that our results may be driven by characteristics typical of all ethno-linguistic enclaves and that there is really nothing specific about our reference points being communities of near-Piedmontese descent. For example, linguistic minorities may display higher social cohesion or have developed internal social structures that reduce their propensity to violently reject transplanted institutions. To address this concern, we replicate our analysis by taking different cultural enclaves (Greek, Croatian and Albanian) as reference points: distance from those enclaves does not affect brigandage incidence in a significant way. This evidence supports the claim that cultural proximity between the environment in which the institutions originated and that in which they are applied plays a crucial role in determining these institutions' ability to function.

After establishing the link between cultural proximity and institutional rejection, we investigate whether there is a relation between our measure of the latter and the levels of political participation in the years following the Italian unification. Specifically, we show that local changes in electoral turnout between the 1861 and 1865 national elections are negatively correlated with the intensity of the occurrence of brigandage episodes between 1862 and 1864.[4] Institutional rejection thus induces lower political participation. We also explore the relationship between brigandage and turnout in subsequent elections. We find suggestive evidence that municipalities with a large number of uprisings had depressed turnout up until the early 20th century. This 40-year-long effect of brigandage suggests that the short-term rejection of Piedmontese institutions impacted the nation building process through its long-lasting influence on political participation.

Our results add new evidence to the literature on institutions and culture.[5] The economic literature has already highlighted how these factors generally evolve jointly (Bisin and Verdier, 2017) and interact with one another (Tabellini, 2010; Alesina and Giuliano, 2015). In particular, while the economic analysis of institutions and institutional change has often implicitly assumed cultural homogeneity in the underlying environment (Acemoglu et al., 2011), empirical evidence suggests that the same institution can have different effects when local norms and social capital are different (Putnam, 1993). For instance, the functioning of legal and administrative institutions and their effectiveness in terms of

[4]This time span was chosen to capture the effect of uprisings occurring between the two electoral rounds: see Section 2.6.

[5]The positive impact of both well-functioning institutions and favorable cultural traits on economic growth has been widely documented in the recent economic literature (see Acemoglu et al., 2005, and Guiso et al., 2006, for general introductions).

economic outcomes are deeply affected by the cultural traits of the environment to which they are applied (see, among others, Acemoglu and Jackson, 2017, Guinnane, 1994 and Guiso et al., 2016.).

The main empirical challenge in identifying the effect of underlying cultural traits on institutional rejection is that institutional variation interplays with cultural variation.[6] Institutional transplantations – the deliberate, rapid and forced exportation of institutions from one location (the *donor environment*) to another (the *recipient environment*) – can provide variation in institutions that is uncorrelated with underlying cultural traits, as the transplanted institutions are almost never designed to be transplanted. Transplantations involving several recipients simultaneously are particularly useful because they often lead to the imposition of a uniform set of institutions on culturally diverse groups.

This is not the first work to make use of institutional transplantations to analyze the relationship between cultural traits and institutional effectiveness. For instance, Lecce and Ogliari (2017) study the long-run economic effect of the interaction between culture and the imposition of French institutions on Prussian counties during the Napoleonic occupation. However, transplanted institutions face the possibility of immediate rejection by the recipient: this rejection may come in different forms, ranging from violent uprising against the donor (as in our case) to the local bureaucracy's prevention of the application of the foreign law or use of the transplanted institutions to different ends than those for which they were originally intended (as in Kurkchiyan, 2009). In this paper, we measure the degree of *institutional acceptance* by recipient communities and investigate whether cultural proximity to the donor facilitates the possibility that the transplanted institutions will take hold in the first place by decreasing the intensity of rejection.[7]

This paper also touches upon the more specific literature on legal transplantation (e.g. Berkowitz et al., 2003*a* and Berkowitz et al., 2003*b*). These studies focus on the long-term economic effects of imported legal institutions and attribute differences in their adoption

[6]For instance, difficulties in studying the interplay between cultural traits and institutions may come from the fact that it is difficult for identical institutions to arise endogenously from different cultural milieus (Tabellini and Greif, 2010); that the same institution may display different degrees of effectiveness when applied to different environments because of the institutional – rather than the cultural – background of such environments (Ma, 2013); and that culture itself may be influenced by institutional arrangements, implying an obvious endogeneity issue (Aldashev et al., 2012). Moreover, cultural traits are difficult to specify and measure in absolute terms: 'culture' is composed of preferences and beliefs, and although several ingenious quantifications have been proposed (e.g., Guiso et al., 2004), all inevitably capture only some aspects of a more general, yet unfathomable, picture.

[7]In a different context, Fisman et al. (2017) stress the importance of cultural proximity in mitigating informational asymmetries in the Indian lending market.

to the lawmaking process and the demand for laws exploiting cross-country variation. We instead concentrate on a single transplantation instance and use finer municipality-level variation to highlight the role of cultural distance from the donor environment in determining short-run institutional rejection.

Finally, we connect our study with other works in the economic literature that focus on Italian post-unitary brigandage. Most notably, Accetturo et al. (2017) exploit a "side effect" of brigandage, i.e., the passage of the draconian Pica Law in 1863, to study the effects of divisive policies on voter turnout in the aftermath of unification. Amodio (2012) shows that three selected and notable brigandage episodes destroyed social capital and had long-lasting effects on voter turnout. To the best of our knowledge, however, our paper is the first to link near-Piedmontese (and, more generally, non-indigenous) ancestry in Southern Italy with the intensity of brigandage.

The remainder of the paper is organized as follows. Section 2.2 reviews the historical background, discussing the situation of the Italian peninsula during the unification process, the guerrilla episodes following the creation of the new state, and the history of the northern cultural enclaves settled in Southern Italy. Section 2.3 illustrates our data and provides descriptive statistics. Section 2.4 presents our identification strategy, illustrates the main results and discusses the relevant mechanisms. Section 2.5 examines a number of alternative channels that may explain our results, while Section 2.6 investigates the long-run relationships between brigandage and social capital. Section 2.7 concludes. Appendix A contains additional figures, tables and robustness checks. Appendix B provides further details on the data and the historical data sources.

2.2 Historical context

2.2.1 Italian unification and brigandage

Our analysis is made possible by the peculiar historical circumstances that arose within the Italian unification process of the 1860s. Between the spring and summer of 1860, after landing in Sicily, general Garibaldi had occupied the territory of the former Kingdom of the Two Sicilies, which included Sicily and continental Southern Italy (henceforth 'Southern Italy') and had hitherto been ruled by the Bourbon dynasty.[8] By the late summer of 1860, Garibaldi ruled as a *pro tempore* dictator in the name of the Savoy after having ousted the legitimate king from Naples, the capital city, and forced him and the remnants

[8]For an administrative map of Southern Italy in the period under study, see Figure 2.4

of the Bourbon army to retreat within the walls of the city-fortress of Gaeta, near the border with the Papal States.[9] Unable to further delay an official military intervention, the Piedmontese army descended into Southern Italy, and by October 1860, the former Bourbon territories had been integrated into the Piedmontese kingdom.

The unification process was carried out by a complete transfer of the Piedmontese institutions onto the annexed territories. The *Statuto Albertino* (the Piedmontese constitution) became the constitution of the newly formed state, and between late 1860 and early 1861, the Piedmontese lieutenants began issuing decrees involving the extension of the whole Piedmontese administrative and judicial system to the rest of Italy and assigned key administrative position to loyal Piedmontese officials. The intention of the Piedmontese to extend their law and administration to the invaded areas was commonly known even before the territorial occupation was completed. Ultimately, the explicit intention of the Savoy king was to unify all Italian-speaking territories under his rule.[10] There was no indication that Victor Emmanuel and Cavour would adopt any form of federalism: following the post-Napoleonic French tradition, Savoy rule was not authoritarian but heavily centralized and aimed at creating a "strong state", disregarding local diversities that could hinder the political unity of the new kingdom.

The transplantation of the Piedmontese institution created a sense of dissatisfaction among some of the elites and most of the common population alike, by whom the *Piedmontization* of the institutions was considered a form of hostile invasion and not of annexation.[11] This sentiment of dissatisfaction provided fertile ground for the Bourbon house to instigate unrest. The first months of Piedmontese rule, in which Bourbon resistance was still thriving within the military strongholds, were characterized by unorganized popular uprisings in rural towns, most of which were sparked by instigators affiliated with either the former king or the Church. This first phase of reaction against the Piedmontese invasion faded as rapidly as the authorities that supported it: Gaeta surrendered in February 1861, and the Church, already deprived of more than half of its former territories, de-

[9]Other southern troops gathered in the strongholds at Messina (Sicily) and Civitella del Tronto (at the northeastern extremity of the kingdom).

[10]The continuity with the earlier Kingdom of Sardinia was also emphasized by two measures of extreme symbolic value: the regnal number of the king – Vittorio Emanuele remained "the second" instead of becoming the "first" king of Italy – and the extension of the Piedmontese currency to the whole kingdom.

[11]In a heartfelt parliamentary motion presented on November 20, 1860, duke Marzio Francesco Proto Carafa Pallavicino, deputy of Casoria, complained about the extensive presence of the Piedmontese in key official positions and clearly stated the sentiment of dissatisfaction felt by the Neapolitan population: "*There is not an instance in which an honest man could earn something without involving a Piedmontese to help. [...] This is invasion, not union, not annexation! This is treating our land as a conquest land. [...] The government of Piedmont wants to treat our land as Cortez and Pizarro did in Peru and Mexico.*"

creased the intensity of its hostilities around that time. Nevertheless, it soon became clear that insurgencies, despite being instigated by loyalist agents, were founded upon deeper roots.

Three aspects of the institutional transplantation deeply resonated with the masses and served as a catalyst for their mobilization. First, in line with its nineteenth-century liberal ideology, the Piedmontese government was unambiguously anti-clerical, and its civil and penal law contained hardline provisions against the Church's temporal power and economic stance. Religious orders were either abolished or deprived of land and other possessions. This constituted a shock to the rural communities of Southern Italy not only on cultural or ideological grounds but also because ecclesiastical organizations would often intervene in favor of the poorest and weakest during times of economic downturn. A similar mechanism underlies the reasons behind the negative reception of Piedmontese land reforms by the lower classes. Southern Italy was organized as a post-feudal economy, with a tiny class of landlords owning large estates (the so-called *latifondi*), an equally small but increasingly wealthy and powerful urban middle class, and a large mass of landless and mostly propertyless peasants.[12] The possessive capacity of those few landlords was limited by the presence of *terre demaniali* – lands directly owned by the Crown that could be used as pasture or farming land by peasants in times of poor harvest and/or low employment (Liberati, 1988).[13] Retrieving an old plan drafted by the Napoleonic government at the beginning of the century (that had already sparked an insurgency at the time: see Pappalardo, 2014), the new government decided to partition and auction off these lands. The reorganization of land usage benefited mainly the bourgeoisie (which supported national unification) but was not well tolerated by the peasants, reinforcing their connection with the former king and the Church, who were considered friendly authorities, as opposed to local nobility and foreign invaders. Finally, the Savoy decrees contained provisions for the inclusion of both ex-Bourbon and new recruits in the newly formed national army: to this end, compulsory military service was planned for 1861 and subsequent years.[14]

[12]In the first decades of the nineteenth century, Cerignola, in northern Apulia, was a wealthier-than-average agricultural town, the development of which was certified by sizable immigration. However, only 40.5% of its inhabitants lived on a larger-than-subsistence income, and 68% of households were headed by landless peasants (Russo, 1988).

[13]To appreciate their extension, consider that, a few decades before national unification, state-owned land constituted 54.4% of the agricultural land in the flat countryside of Cerignola (Russo, 1988) and 39% of the mountainous municipality of Morcone (De Francesco, 1988).

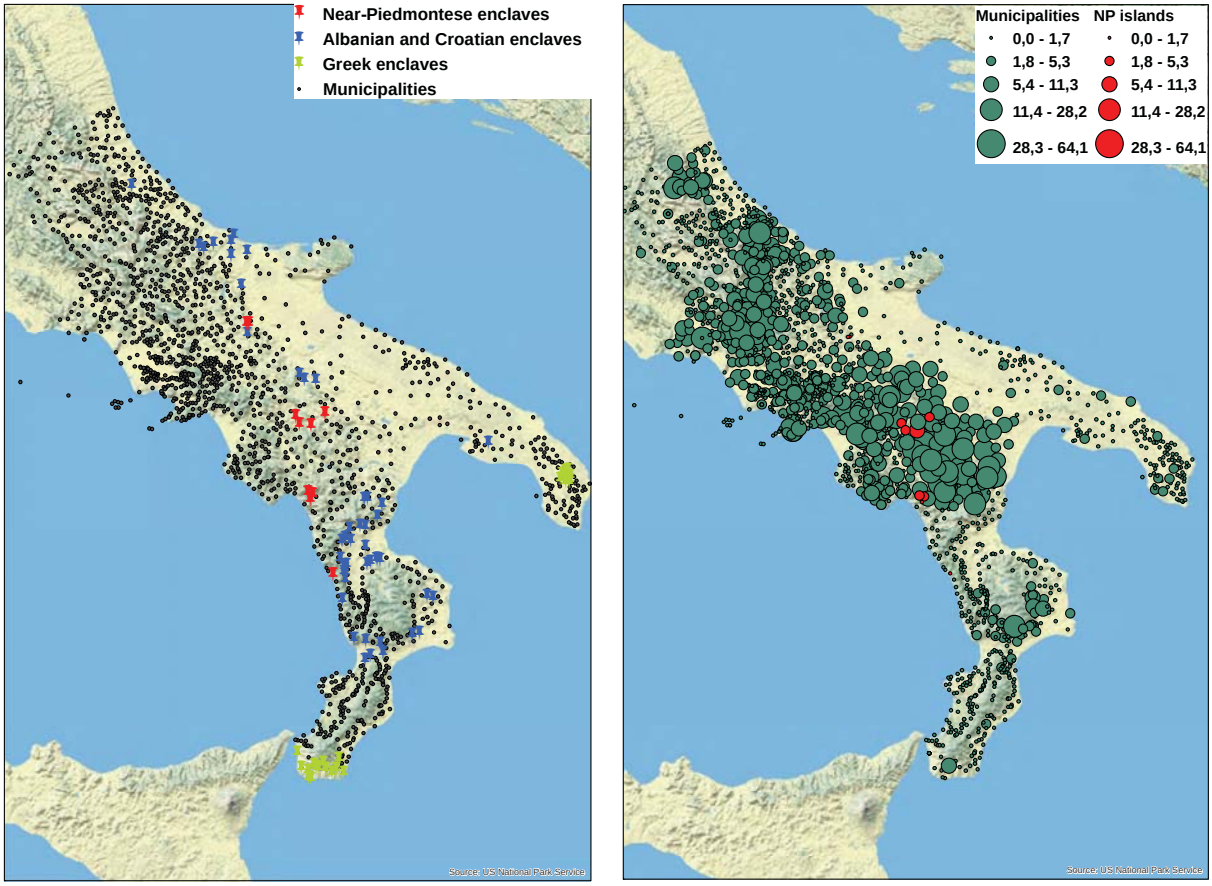
[14]The unraveling of Piedmontese plans for the re-organization of the Southern army epitomizes the popular opposition to the regime change: a decree was issued on December 20, 1860, to reintegrate 70,000 Bourbon soldiers into service under Piedmontese command. The deadline was delayed several

Thus, by 1861 all the unwelcome innovations coupled with the demise of the reference authorities of previous centuries (the Bourbon king and the Church) sparked a new wave of popular unrest that – in some communities – led to the organization of guerrilla groups of so-called *brigands*. Brigands came almost invariably from the peasant class and were joined in some cases by disbanded Bourbon soldiers who refused to integrate into the new army's ranks. Their bands varied in size, ranging from few individuals to hundreds: one of the best-known brigand leaders, Carmine Crocco, claimed to have once led an army of more than 2,000 (Ciocca, 2013). Bands existed almost exclusively in rural areas, and due to their need to escape regular troops, were concentrated where control of the territory by the non-indigenous was more difficult, i.e., around the mountain range of the Apennines and the hillside areas to the north and east of Naples (Figure 2.1).

Scholars partition the brief history of brigandage into three phases. The first coincides with the formation of brigand bands throughout the years 1860-1861. The influence of the Church and the king and the hope that the latter would return – and old institutions thus restored – were the main drivers of brigand activity in this period. This activity involved instigating and leading episodic revolts in rural municipalities; attacking non-loyalist landowners, especially the urban elites who had acquired lands thanks to the Piedmontese reforms; and, occasionally, clashing with regular troops. Between incursions, brigands retreated to secluded areas that they left only to obtain supplies, an activity that was constantly supported by the local peasantry. The second phase saw brigand groups organizing in paramilitary fashion, with commanders-in-chief and a more stable, albeit fragmented, organization. Despite Piedmontese efforts,[15] intense guerrilla warfare lasted until 1864. Extending a label used by historians, we call these first two phases 'great brigandage'. Unable to cope with such political instability in half of the national territory and facing a seemingly endless drainage of resources, the government introduced a form of martial law (the 'Pica law') in August 1863, with the goal of providing "temporary and exceptional means of defense" for public order. The hardening of legal provisions (going as far as allowing the execution of suspects without a trial) and the near-absolute power given to the military in 11 out of 16 provinces (those labeled as "infested by brigands") in Southern Italy were effective, and the intensity of brigand activity began to fade in the second half of 1863. Beginning in 1865, political brigandage began gradually losing the connotation of anti-Piedmontese resistance and progressively faded into common criminality.

times, but by June 1861, only 20,000 had appeared for service.

[15] Apparently, at the height of the anti-brigandage operations in early 1864, the Italian army had deployed more than 110,000 soldiers against the brigands, amounting to approximately two-thirds of the available units (Ciocca, 2013).



(a) Near-Piedmontese communities (red pinpoints) and other linguistic enclaves (blue and green pinpoints).

(b) Intensity of brigandage (in episodes per 1,000 inhabitants). Near-Piedmontese communities are denoted by the red dots.

Figure 2.1: Municipalities in continental Southern Italy as of 1861.

Post-unitary brigandage was an explicit, violent form of reaction against the institutions transplanted by the Piedmontese and, in particular, of those provisions of the law that directly interfered with the life of common peasants.[16] This brigandage was a large-scale phenomenon in Southern Italy: a true civil war. According to official data reported by

[16]The anti-institutional sentiment and the objective to safeguard the poor masses are clear in a brigand song from Mercato Cilento: "*Tu si' lu giurici re li miei signuri, i' so' lo capo re li fuorilegge; tu scrivi co' la penna e dai ruluri, i' vao ppe' lu munno senza legge. Tu tieni carta, penna e calamaio ppe' castia' a sti poveri pezzienti, i' tengo povole e chiummo, quanno sparo: giustizia fazzo a chi non tene nienti.* (Translation: You are the judge of the lords, I am the head of the outlaws; you write by pen and cause sorrow, I go around the world without law. You have paper, pen and inkpot to punish these poor souls, I have got gunpowder and bullets, when I shoot: justice I make for those who have nothing.)"

Molfese (1964), of a total number of brigands estimated at approximately 80,000, more than 5,000 had been killed (during military operations or by execution) between 1861 and 1865, a similar number had been arrested and approximately 3,600 had surrendered. According to Ciocca (2013), approximately 6,500 brigands and more than 1,600 regular soldiers were killed in the more extended period of 1861-1869. In this paper, we use the number of brigandage-related events in each municipality of Southern Italy as a measure of the intensity of the rejection of the institutional transplantation during the Piedmontese invasion of Southern Italy and the consequent process of national unification.

2.2.2 Near-Piedmontese enclaves in Southern Italy

The aim of this paper is to test the hypothesis that cultural traits matter at least partly for institutional acceptance. We focus on a specific aspect of the cultural identity of the receiving environment, namely, its cultural proximity to the donor. Our task is therefore to test whether cultural proximity has a (positive) impact on institutional acceptance. To do this, we rely on the existence of communities of near-Piedmontese descent in areas where brigandage was, on average, intense.

Origins

There is substantial agreement among historians that such communities were established in the late Middle Ages (with the earliest mentions dating back to between the thirteenth and fifteenth centuries). We call them 'near-Piedmontese' because not all of the places of origin of the immigrants who first settled these communities lie within the current borders of Piedmont (some trace their origin back to the contemporary region of Liguria; see Toso, 2002) or within the 1861 borders of the Kingdom of Savoy (other communities are thought to be descendants of soldiers from the Alpine valleys of what is now southeastern France; see De Salvio, 1908). What matters for this paper is that all these communities originated within the Provence-Savoy-Piedmont area, which is also the cradle of the Kingdom of Savoy and the macro-region where its cultural traits developed. Ten near-Piedmontese communities are easily identifiable (and were at the time of the Italian unification) because they retain Gallo-Romance dialects.[17] For most such municipalities, the Gallo-Romance dialect was the only language spoken by peasants at the time of the events we study, although it is believed that most could communicate, at least at a basic level, in the

[17]We use an extensive definition of 'Gallo-Romance languages' that encompasses Franco-Provençal, Occitan and the Gallo-Italic languages of Piedmont and Liguria.

language of the surrounding provinces, and many would have understood Standard Italian because of partial mutual intelligibility (De Mauro, 1963).

Linguistics helps us to identify the places of origin and, consequently, to clarify the reasons underlying the migratory phenomena that led to the formation of these communities. The ten near-Piedmontese enclaves constitute four geographically distinct clusters: the municipalities of Celle di San Vito and Faeto in northern Apulia; seven municipalities in Basilicata, divided into two clusters (Picerno, Pignola, Tito and Vaglio on one hand and Nemoli, Rivello and Trecchina on the other); and the isolated municipality of Guardia Piemontese in northern Calabria. Despite the scarcity of documents, contemporary historians tend to exclude the possibility that the main motive for these communities settling in Southern Italy may have been an attempt to escape religious persecution. The origin of the Apulian Franco-Provençal settlement has been traced back to soldier relocation and land assignment by Charles of Anjou in the 1260-1270s (see, for instance, De Salvio, 1908 and Melillo, 1959). Based on linguistic evidence, Pfister (1991) and Toso (2002) suggest that most of the Gallo-Italic speakers from Basilicata descend from southern Piedmontese and/or Ligurian colonizers, who are not believed to have been interested in religious struggles. The Calabrian settlement of Guardia was indeed created by a Waldensian community, and it is possible that southward migration was sparked by fears of persecution following the Albigensian Crusade (early thirteenth century). There is no evidence, however, that religious motives affected the choice of the precise location of the colony (which might have been otherwise explicitly selected for its secludedness or defensibility). In fact, the Waldensians were positively received by the Calabrian nobility and faced no difficulties until after the Reformation when, because of their adherence to Protestantism, they suffered harsh persecution (Vegezzi Ruscalla, 1862).

Cultural persistence

It is crucial for our empirical strategy that these ethno-linguistic enclaves retained their ancestors' cultural traits over the centuries.[18] Historical evidence documents a persistent cultural similarity between the inhabitants of these near-Piedmontese enclaves and their ancestors. This similarity is made particularly evident by the survival of ancestral linguistic traits, which is usually associated with the persistence of other, less evident cultural traits.[19] German linguist Gerhard Rohlfs was the first to study the peculiar

[18] The persistence of cultural traits has been widely documented in the economic literature. See for example Voigtländer and Voth (2012), Doepke and Zilibotti (2008) and Alesina et al. (2013).

[19] A growing empirical literature studies the relationships among ancestry, language diffusion and cultural persistence over time and space; see Spolaore and Wacziarg (2016*a*) for further details.

linguistic traits of near-Piedmontese municipalities in Basilicata. He reported how local dialects displayed several analogies with Gallo-Romance languages (including the Sicilian Gallo-Romance dialect) and structurally differed from the dialect spoken in the surrounding region.[20] Linguistic similarities are also reported for the other three enclaves. Vegezzi Ruscalla (1862) provides an ethnographic study of Guardia Piemontese in which he thoroughly reports the analogies between the local language and that spoken in the Piedmontese valleys where the ancestors of Guardia's inhabitants originated and underlines the persistent resemblances with the Piedmontese.[21] Finally, De Salvio (1908) led a comparative study of the dialect spoken in Celle and Faeto at the beginning of the twentieth century and the old Savoy dialect of the Middle Ages: the results of this comparison proved that these languages were closely connected. In addition to their linguistic traits, near-Piedmontese communities have retained other aspects of their ancestral culture: writing around the time of our events of interest, Vegezzi Ruscalla (1862) reports that the inhabitants of Guardia Piemontese maintained very similar customs to their Piedmontese ancestors in terms of clothing and, more important for our analysis, attitudes toward property and work.[22] We thus exploit these linguistic and cultural enclaves to proxy for the cultural distance of Southern Italy's communities from the donor environment, assuming that near-Piedmontese enclaves embody customs and social norms very similar to those prevailing in Piedmont. Importantly for our analysis, though, historical accounts provide no evidence of such communities – with the possible exception of Guardia – being recognized as culturally similar by the invading Piedmontese of the time. For example, while the presence of Albanian and Greek communities was reported in the 1861 Census, near-Piedmontese enclaves did not receive any mention in the official statistics. This fact also suggests that Piedmontese officials were likely unaware of their ancient link with these communities. This suggestion arguably excludes the possibility

[20]Rohlfs identifies the core of a northernmost near-Piedmontese cluster in the municipalities of Potenza, Tito, Picerno, Pignola and Vaglio. Moreover, he reports Trecchina (the municipality with the strongest resemblance to northern languages), Rivello, Nemoli and San Costantino (currently part of the municipality of Rivello) as part of a southern Gallo-Romance cluster: Figure 2.8 in Appendix A depicts a map of the two near-Piedmontese clusters in Basilicata. Figure 2.7 in Appendix A is a comparative table of the languages spoken in the near-Piedmontese community of Tito in San Chirico (a nearby non near-Piedmontese community) and in two municipalities in Liguria and Piedmont (the areas of origin of near-Piedmontese colonizers). See Rohlfs (1988) for further details.

[21]Figure 2.9 in Appendix A offers several examples of linguistic similarities and differences comparing the Italian language, the dialect spoken in Guardia Piemontese, the dialect spoken in Val d'Angrogna and the dialect spoken in Cosenza (the largest city near Guardia Piemontese).

[22]Rohlfs (1972) also reports strong similarities in clothing between inhabitants of Guardia Piemontese in the first half of the nineteenth century and their Piedmontese ancestors.

that the Piedmontese were in the position to apply any kind of 'preferential treatment' to near-Piedmontese communities during the period of institutional transplantation.

2.3 Data and variables

The main dependent variable in our analysis is the intensity of brigandage, which we measure as either the total number of brigandage-related episodes in each municipality or as brigandage incidence (i.e., the number of episodes divided by the municipality's population, measured in thousands of units). We use these measures of brigand activity to quantify the intensity of the rejection of the institutional transplantation in each of the 1,855 municipalities in post-unitary Southern Italy (see Figure 2.1). We digitized the information collected in three volumes published by the Italian Ministry of Culture,[23] which include all episodes of brigandage reported in the State Archives of Southern provinces (for an example of entries in the State Archives, see Figure 2.6). This process led to a total of 12,242 brigandage episodes for which we know the geographical location, the type of offense, the authority that reported it and the year in which it was recorded. On average, the municipalities in our population experienced approximately 6.6 episodes each (corresponding to approximately 1.93 episodes per thousand inhabitants), with approximately 68.5% of them experiencing at least one episode and the most intensely brigand-stricken municipality reaching 64 episodes per thousand inhabitants.[24] Based on information contained in the original records, we are able to classify episodes into four broad categories of offenses: violent crimes, clashes with authorities and armed insurrections (36.3%); non-violent crimes such as theft, arson and instances of connivance (36.8%); and, finally, a residual category including all episodes reported as notifications of the presence of brigand bands, reports by citizens or arrests (26.8%). Furthermore, just under half of our episodes (45%) were recorded by courts at any level; most of the other half were recorded by public safety institutions such as the police or the *Prefetture*, the local representatives of the Ministry of the Interior (46%).[25] Approximately one in four (24.1%) occurrences belong to the early phase of brigandage (1860-1861), characterized by the immediate reaction to Piedmontese occupation; approximately three-quarters of our episodes (74.3%) are asso-

[23] Ministero per i Beni e le Attività Culturali (1999-2001).

[24] The total number of episodes exceeds that indicated by most historians who focus on major brigandage events. For example, the clashes reported by Molfese (1964) include approximately 475 major episodes concentrated in approximately 19% of the municipalities in our sample.

[25] Since the records come from different sources, there could be over-reporting of some episodes. We address this issue in Section 2.4.3 and in Appendix A by considering only episodes reported by the police and re-aggregating observations, respectively, to eliminate most potential duplicates.

ciated with the most intense phase of brigand guerrilla war (1860-1864); the remaining 25.7% were recorded between 1865 and 1870.[26]

The main explanatory variable is the logarithm of the geographic distance of each municipality from the closest near-Piedmontese community, which we measure in two different ways. In the baseline specifications, we use the linear distance between municipal centroids. One might argue that because culture typically spreads through frequent contacts and social interactions, for the matter at hand, a measure of actual traveling distance between two places may be more appropriate than linear distance. For this reason, in a subsequent specification, we compute distances using the ancient Roman road network as reconstructed by McCormick et al. (2013).[27] The two measures display, as expected, a highly positive degree of correlation (0.94).

In addition to geographic characteristics, we control for socio-demographic and economic features of Southern Italy's municipalities and provinces before unification. For this purpose, we collect and digitize novel data from several statistical sources dating back to the last decades of the Kingdom of Two Sicilies,[28] and population data concerning the years immediately following Italian unification, obtained from the Italian Census of 1861 and the additional statistical reports of the following decade. To capture differentials in economic growth across municipalities prior to the onset of the events we study, we compute the population growth rate between the 1850s and 1861 (both values are reported in the Italian 1861 Census). Other variables that were recorded before national unification (most of which between 1830 and 1850) at the municipal level include indicators for the presence in each municipality of civil, criminal or commercial courts; of the local episcopal or archiepiscopal seat; of secondary education institutes;[29] of new hospitals; and of relevant manufactures or proto-industrial plants. We also use information reported

[26]When considering the temporal distribution of brigandage, however, one must bear in mind that different institutions might have recorded episodes with varying delay: for instance, courts may have recorded some episodes at the beginning of the associated trials, which may have occurred months or even years after the suspected crime had taken place. In particular, several episodes located in the last phase might in fact refer to brigandage activities that occurred before 1865.

[27]The choice of using a log-transformed version of these distances is in line with our interpretation of physical distance as an inverse proxy for cultural proximity: the diffusion of culture from a point of origin suggests that equal increases in physical distance should be deemed less important as their distance from the origin increases.

[28]Our main data sources are the *Annali Civili*, an official statistical publication by the Ministry of the Interior of the Kingdom of Two Sicilies, published in several volumes between 1833 and 1860; the statistical collection (*Statistica dell'Italia*) by Count L. Serristori published in 1842; and historical works on the sale of national wealth during the Napoleonic period (Villani, 1964) and on the reaction to provisions of land redistribution during the nineteenth century (Corona, 1995).

[29]These are divided into *licei*, *collegi* and other secondary schools.

by Corona (1995) to construct an indicator of popular attitudes toward innovation and the privatization/distribution of common-use lands[30]. We further collect information on provincial-level pre-unitary characteristics: the number of individuals subject to military draft in 1834; the total number of landowners, farmers and fishermen; and the total number of professionals, artisans and servants. We also include provincial data on tax revenues, municipal revenues and total expenditure for public works in 1850-51. Moreover, using data collected by Villani (1964), we can obtain some provincial measures of the intensity and patterns of the diffusion of monasteries and the sale of their lands during the Napoleonic period.

Table 2.1: Descriptive statistics and means comparison

	nP communities (1)	Other municip. (2)	Difference (3)
Area 1861 (ha)	4361.30	4163.67	-197.63
Pop. 1861 (ths.)	3.01	3.66	0.65
Pop. growth	-0.11	-0.04	0.06
Altitude (100m)	6.71	4.14	-2.58***
Dist. Naples	138.98	160.19	21.20*
Dist. n. prov. seat.	33.64	32.13	-1.51
Dist. town	30.59	19.01	-11.58
Dist. coast	35.17	22.11	-13.06
Dist. Papal States	218.56	191.57	-26.99
Observations	10	1845	1855

Column 1 reports means in near-Piedmontese communities; column 2 reports means in all other municipalities; column 3 reports the difference between the two means: tests for the difference in means allow for unequal variances in the sub-samples (* p<0.1, ** p<0.05, *** p<0.01). All distances are expressed in kilometers. For a detailed description of the variables, see Appendix B.

Table 2.1 reports the mean values of several geographic controls for near-Piedmontese communities and all other municipalities in Southern Italy. Differences in such values are statistically significant only in the case of altitude and distance from Naples, with nP communities lying on higher grounds and closer to the capital than the average municipality. Consistent with this evidence, we notice from Table 2.2 that near-Piedmontese communities display slightly more intense brigandage than the average municipality. However,

[30]The information provided by Corona (1995) refers to the last decades of the eighteenth century and therefore predate the period under analysis by 70-80 years. We provide more information on the data and on the index we construct in Section 2.5.1.

when the comparison population is reduced to the subsample of municipalities within the first quartile of distance from the closest near-Piedmontese community, we see that the latter display much less intense brigandage. In particular, it emerges from Figure 2.2 that most municipalities within this restricted range display more brigandage than the national average; that the incidence is smaller for the 10 near-Piedmontese communities than for their most immediate neighbors; and, finally, that there is an evident decrease in incidence as one moves closer to near-Piedmontese communities within a 25 km radius.

Table 2.2: Brigandage intensity by distance from the nearest nP community

	All municipalities			nP communities			Dist. 1st quartile		
	mean	min	max	mean	min	max	mean	min	max
Episodes (tot.)	6.60	0	211	8.80	0	26	11.21	0	211
Episodes p.m.	1.93	0.00	64.13	2.33	0.00	6.59	3.04	0.00	64.13
Observations		1,855			10			461	

Episodes (tot.) is the number of brigandage episodes in each municipality; while *Episodes (p.m.)* is the number of episodes per 1,000 inhabitants (a measure of brigandage *intensity*) in each municipality. Column 1 reports means, minimums and maximums for all municipalities; column 2 reports statistics for near-Piedmontese communities only; column 3 reports statistics for non-nP municipalities whose distance from the nearest nP community lies in the first quartile of such distance's distribution in our sample. These correspond to all non-nP municipalities within a 46 km radius from the nearest nP community. For a detailed description of the variables, see Appendix B.

The fact that near-Piedmontese communities display above-average brigandage incidence should come as no surprise. As it emerges visually from Figure 2.1b, the majority of brigand activities were concentrated along the Apennine range and, in general, in inland municipalities rather than in coastal towns. Near-Piedmontese communities meet these characteristics.

Table 2.3 reports the average of our main geographic and socio-economic controls for municipalities with a brigandage incidence below the median (column 1); for the complementary set of municipalities with incidences above the median (column 2); and, finally, for the municipalities most affected by brigandage (those with incidences in the top quartile, column 3). The patterns are as expected: communities with higher levels of brigandage tend to be larger in area but smaller in population, consistent with their location at the center of the peninsula, in mountainous areas, relatively far from the coast and major towns. The distance from Naples, the kingdom's capital city, appears to be negatively correlated with brigandage intensity, as both areas near Naples are attractive to brigands because of their wealth (the province of Naples exhibited, for instance, a larger population

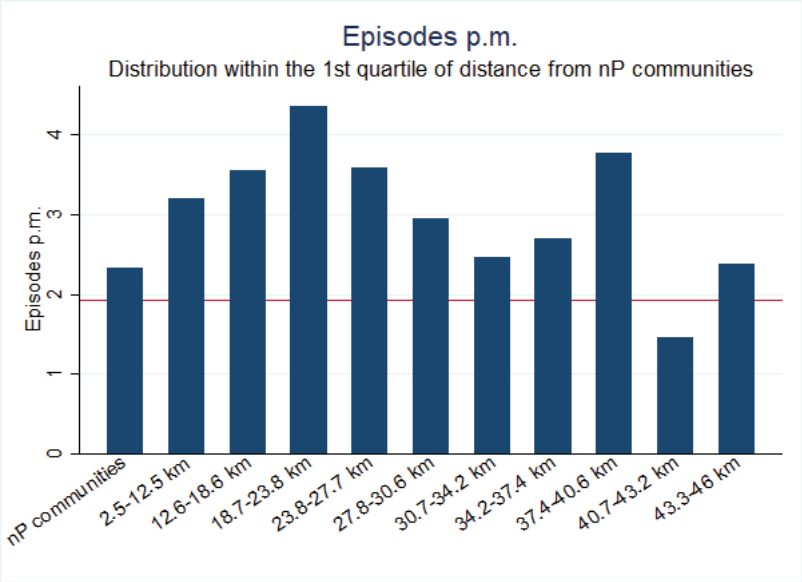


Figure 2.2: Distribution of episodes per thousand inhabitants within the first quartile of the distance from the nearest near-Piedmontese community (46 km). Each bin includes 46 or 47 municipalities; labels on the horizontal axis indicate the distance range of each bin. The horizontal red line denotes the population average (1.93).

growth rate than the rest of Southern Italy over the 1850s) and because brigandage was almost absent from the southernmost extremities of Southern Italy, i.e., the tips of Apulia and Calabria, which are the areas located farthest from Naples. The same geographic disposition explains part of the evidence on the distance from the Papal States (lying beyond the northernmost border of the kingdom), to which the explicit support given to brigands by the Church in the earliest phases may have also contributed.

2.4 Cultural distance and institutional rejection

In this section, we present our main results. As described in Section 2.2, Italian unification was a common 'shock' that generated a reaction with different degrees of intensity across Southern Italy. In some municipalities, citizens responded with violent uprising and riots, while in others, no episodes of protest were reported. We argue that geographical distance from a community of near-Piedmontese ancestry is a significant determinant of this institutional rejection, which is evidence that cultural distance from the donor environment hinders institutional acceptance. Indeed, geographical distance represents an inverse measure of the likelihood of historical interaction with near-Piedmontese communities and the associated cultural admixture, providing a good indicator of cultural

Table 2.3: Descriptive statistics by brigandage intensity

	Occasional brig. (1)	Frequent brig. (2)	Highly freq. brig. (3)
Area 1861 (ha)	3752.84	4577.08	5054.14
Pop. 1861 (ths.)	3.72	3.59	3.48
Pop. growth 1824-61	-0.06	-0.03	0.00
Altitude (100m)	3.77	4.53	5.09
Dist. Naples	192.03	128.08	115.60
Dist. n. prov. seat.	31.44	32.84	35.29
Dist. town	18.04	20.11	22.10
Dist. coast	19.22	25.14	28.03
Dist. Papal States	227.31	156.07	144.30
Dist. Piedm.	98.92	78.61	69.88
Observations	928	927	463

All distances are expressed in kilometers. Municipalities are classified according to brigandage intensity. *Occasional brigandage* denotes municipalities with less than the median value of episodes per capita (.725); *Frequent brigandage* denotes municipalities with more than the median value of episodes per capita; *Highly frequent brigandage* denotes municipalities with more than the 75th percentile of episodes per capita (2.35). For a detailed description of the variables, see Appendix B.

proximity to the origin of the institutional arrangement. We show that after controlling for a number of geographic and socio-economic observables, the intensity of brigandage is positively correlated with the distance to near-Piedmontese communities.

2.4.1 Comparison of two communities: a case study

To clarify our point, we consider as an example two villages that are comparable in terms of inhabitants and geographical characteristics. Let us consider Castelluccio Superiore, a municipality in Basilicata with a population of 2,900 in 1861 and an altitude of 680 meters, and Castelluccio Valmaggiore, an Apulian municipality with a population of approximately 2,700 and an altitude of 630 meters (the similarity in names is accidental: four municipalities in total share the denomination of 'Castelluccio'). The former was located 16 km from one of the two near-Piedmontese clusters in Basilicata, with the closest enclave being Nemoli. The latter was 2.5 km away from the near-Piedmontese municipality of Celle San Vito. Many brigandage episodes were reported in the area of Castelluccio Superiore (24 episodes, with an incidence of 8.3 episodes p.m.). Molfese (1964) mentions Castelluccio Superiore in his book, describing the "assault of Castelluccio" on August

1863, when 20 national guards were assaulted by 40 brigands while escorting a group of noblemen. In the ensuing clash, nine people (six guards, one noblemen and two brigands) died, and the surviving noblemen were kidnapped. Castelluccio Valmaggiore provides a stark contrast with Castelluccio Superiore. No episodes of brigandage were reported in our sources during the military invasion or in the following years. However, Daunia, the broader area in which Castelluccio Valmaggiore lies, was not immune from brigandage: Lucera, its main center lying 24 km from Castelluccio Valmaggiore, was an intensely brigand-stricken municipality. We now investigate whether this observation can be generalized to claim that proximity to a near-Piedmontese community was associated with less-intense brigandage, conditional on the observables.

2.4.2 Empirical strategy

To test our central hypothesis, namely, that *ceteris paribus*, Piedmontese institutions were more harshly rejected in regions farther from near-Piedmontese communities and therefore less exposed to the influence of a near-Piedmontese culture, we estimate several generalized linear models of the following form

$$g\left(\mathbb{E}(Y_{i,j} \mid \mathbf{X}_{i,j})\right) = \beta_1 \text{Dist}_i + \beta_2 \text{Pop}_i + \beta_3' \mathbf{G}_i + \beta_4' \mathbf{C}_i + \beta_5' \text{Prov}_j \tag{2.1}$$

where $Y_{i,j}$ is the number of brigandage episodes recorded in town i of province j or, in some specifications, the incidence of brigandage in town i of province j; $\mathbf{X}_{i,j}$ denotes all regressors related to town i of province j, which include Dist_i, a measure of the distance of municipality i from the closest near-Piedmontese community, our main independent variable; Pop_i (a suitable transformation of the) population of town i;[31] $\mathbf{G}_i$ and $\mathbf{C}_i$, vectors of geographical and pre-unitary socio-economic controls, respectively, measured at the municipality level; and Prov_j, a set of pre-unitary controls measured at the province level and possibly reduced to a set of province indicators. The key coefficient is β_1, the effect of the distance from the closest near-Piedmontese community on the number of brigandage episodes or the incidence of brigandage. We expect β_1 to be positive and significantly different from zero, indicating that – conditional on our controls – proximity to

[31] When g is not linear and the specification requires the dependent variable to take on integer values, the impact of regressors on the incidence of brigandage, which is measured in linear models as the number of episodes per 1,000 inhabitants, is found by retaining the number of episodes as the dependent variable and forcing $\beta_2 = 1$ using log-transformed population as Pop_i. In this case, as β_1 is interpreted as the effect of distance on expected incidence passed through g, we shall indicate brigandage incidence as the dependent variable in our tables.

a near-Piedmontese community reduces a municipality's propensity to experience brigand activity in its territory.

Our identification relies on the absence of correlation between our measure of distance and the error terms, conditional on all controls. Taking settlement locations as fixed, this amounts to claiming that the location of near-Piedmontese communities must be randomly determined. Among other factors, we control for time-invariant observables, such as altitude and the linear distances from the coastline and the capital city, which, as presented in the previous subsection, make the assignment as a near-Piedmontese community approximately random. Moreover, while near-Piedmontese communities are the result of migratory movements, historical evidence presented in Section 2.2 suggests that the first colonizers did not autonomously decide where to settle, having been either invited (by landowners) or ordered (by political and military authorities) to take possession of those areas. Thus, the location and characteristics of near-Piedmontese communities were not chosen according to criteria that might, through other channels, influence the presence and/or intensity of brigand activity.

Our interpretation of the results relies on the dual assumption that culture exhibits long-run persistence and, simultaneously, that the near-Piedmontese communities on which we focus had interacted with their immediate neighbors in previous centuries. The first claim – discussed in Section 2.2.2 – guarantees that these communities retained traits of their ancestral culture within a homogeneous and culturally distant surrounding environment, from their foundation until the events under study. The second claim – which we examine in depth in Section 2.4.4 – justifies our use of geographical distances as proxy for cultural distance, as social interaction must have implied that neighboring communities came into contact with the near-Piedmontese cultural heritage.

2.4.3 Main results

Table 2.4 reports estimates of ordinary least squares models (columns 1–3) and negative binomial models (columns 4–6) that relate the total number of brigandage episodes to the *linear* distance of each municipality from the closest near-Piedmontese community and a varying set of controls.[32] The relevant distances are log-transformed.[33] Columns 7–9 report estimates of ordinary least squares models using the incidence of brigandage (i.e., the total number of episodes is normalized by the population in 1861) as the dependent

[32]Tables 2.11 and 2.12 in the Appendix report all the estimates omitted from Table 2.4.

[33]In operating this transformation, we set distances in the interval $[0, 1)$, which includes only the zero-distance of near-Piedmontese communities from themselves, to 0.

variable. Columns 10–12 report the result for a negative binomial model in which the restriction $\beta_2 = 1$ applies (see footnote 31). Columns 1 and 4 include the most basic demographic controls: population growth in the 1850s-1861 period (which may also serve as an indicator of economic growth in this pre-industrial setting), the log-transformed municipal population as reported in the 1861 Census and the main geographical controls (municipal coordinates, area and altitude), distance controls, provincial controls and a set of region-level fixed effects. Columns 2 and 5 replace regional fixed effects with province-level fixed effects, which also absorb all the provincial controls, while columns 3 and 6 complete the specifications by adding pre-unification municipal-level variables. Columns 7–9 do not include the log-transformed municipal population when using the linear model in order to match the aforementioned restriction on the coefficient for the population in count models 10–12. The most complete specifications (namely, models 3, 6, 9 and 12) all display a positive and significant correlation between the intensity of brigandage and the distance from the nearest near-Piedmontese community.

Clearly, we do not claim such distance to be the only driver of brigandage intensity, but its effect is sizable: according to the estimates in model 9, doubling the distance from the nearest near-Piedmontese community is associated with a municipality experiencing 0.21 additional episodes p.m. (amounting to an increase of 10.5% on an average of 1.93 episodes p.m.); model 12 suggests that doubling the distance is associated with a 17% increase in the expected incidence.

An inspection of the other coefficients, reported in Tables 2.11-2, shows the resilient relevance of the municipality's distance from Naples, which is associated with a decrease in brigandage. Population and area, which are trivially positively correlated with the number of episodes, nonetheless appear to influence the incidence of brigandage in opposite directions: geographically larger municipalities tend to display a higher brigandage incidence, but – as we noted above – the latter decreases with the municipality's population. As anticipated, brigandage is associated with higher altitudes and a longer distance from the nearest coast.

Results reported in Table 2.5 are meant to assess the stability of our baseline findings and to clarify the interpretation of the estimates. As explained previously, brigandage displayed the most reactionary and insurrectional features in the early phases of its existence before transforming into a less politically motivated – albeit widespread – criminal movement.

Column 1 of Table 2.5 replicates the baseline specification (i.e., model 12 of Table 2.4) using only episodes occurring between the years 1860 and 1864, a period which – extending the historians' definitions – we label 'great brigandage'. In column 2, we exploit

Table 2.4: Baseline results

Dep. Variable:	Total Number of Episodes (Ep. tot)						Episodes per thousand inhabitants (Ep. p.m.)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(log) Dist. Piedm.	1.796*	1.752*	1.517*	0.185**	0.178**	0.169*	0.717**	0.706**	0.699**	0.225**	0.222**	0.227***
	(0.937)	(0.944)	(0.860)	(0.091)	(0.090)	(0.087)	(0.290)	(0.292)	(0.289)	(0.090)	(0.088)	(0.086)
Model	OLS	OLS	OLS	Neg. Bin.	Neg. Bin.	Neg. Bin.	OLS	OLS	OLS	Neg. Bin.	Neg. Bin.	Neg. Bin.
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Distance controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Provincial controls	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No
Municipal controls	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Region FEs	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No
Province FEs	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Adj./Ps. R^2	0.422	0.423	0.488	0.155	0.157	0.163	0.352	0.354	0.353	0.111	0.113	0.118
Obs.	1855	1855	1855	1855	1855	1855	1855	1855	1855	1855	1855	1855

Pop. growth included in all specifications. Columns 1-6 also include (log) Pop. 1861; in columns 10-12, Pop. 1861 is used as the exposure variable. Geographical controls include: (log) Area 1861 (ha), Altitude (100 ms), Latitude and Longitude. Distance controls include: (log) Dist. Naples, (log) Dist. n. prov. seat, (log) Dist. town, (log) Dist. coast, (log) Dist. Papal States. Prov. controls include young men, real estate owners, liberal arts professionals, farmers, artisans, fishermen, average duties on milled grain, state revenues, public expenditure, monasteries suppressed in 1806-1815, monasteries reinstated in 1818, buyer per contract and rent per contract in the alienation of national wealth occurred in 1806-1815; all variables refer to pre-unitary years. Municipal controls include dummy variables for the presence in each municipality of civil, criminal or commercial courts; of the local episcopal or archiepiscopal seat; of secondary education institutes; of hospitals; and of relevant manufactures or proto-industrial plants. Standard errors clustered at district level in parentheses (* $p<0.1$, ** $p<0.05$, *** $p<0.01$).For a detailed description of the variables, see Appendix B.

information on the types of crime reported in the archives. Although it is informative to consider the whole set of brigandage episodes together, it is essential to test that our main hypothesis still holds for the subset of events with the most vivid anti-establishment connotation. In particular, we restrict our analysis to episodes that either are reported as having explicit political content (e.g., the organization of popular insurrections or clashes with regular police or army forces) or episodes that are classified as violent in nature: these account for approximately 36% of all brigandage episodes in our records. The estimated coefficients remain similar in both magnitude and significance to those found in our previous results. As explained in our description of the data, observed brigandage episodes were recorded by several sources. Typically, police and other public safety forces would record episodes as soon as they were reported to them by officers or citizens. We deem these sources as more reliable both in terms of geographical localization and in terms of temporal positioning. In column 3, we therefore report estimates of our main specification using brigandage data recorded by police forces only. This exercise also allows us to dispel any doubt about the existence of multiple entries for the same episode across different sources.[34] In column 4, we propose an alternative measure of geographical proximity. As social interaction depends on suitable means of communication, we found it appropriate to check whether our results held when computing road distances rather than linear distances. This might be particularly relevant, given the complex orographic features of the areas we study. Given historical evidence on the high degree of coincidence between nineteenth-century roads and the ancient network of Roman roads, we chose to use the latter (McCormick et al., 2013) to compute our approximate road distances. Indeed, the ancient Roman road network was the main communication system throughout the Middle and Modern Ages, i.e., the centuries in which cultural traits from the near-Piedmontese communities might have spread to neighboring communities through socio-economic contact (see Figure 2.5). Although other minor roads not included in our network are likely to have been in used in these centuries, there is evidence that, well into the nineteenth century, many of the roadways in use in Southern Italy still exploited the ancient Roman tracks. The effect that we obtain is extremely similar to that derived using linear distances.

In Appendix A, we perform an extensive sensitivity analysis addressing possible measurement and specification issues. First, we show that the results are robust to the use of alternative measures of brigandage intensity. In particular, we estimate our main specification using brigandage data reported by courts and by other sources, and we recode our dependent variable aggregating all entries that share the same date and the same place

[34]Results using episodes reported by other sources are provided in Table 2.13 in Appendix A.

Table 2.5: Refinements to baseline results

Dep. Variable: **Ep. p.m.**	Great brigandage (1)	Political or violent (2)	Reported by police (3)	Roman roads (4)
(log) Dist. Piedm.	0.229**	0.204***	0.214**	
	(0.113)	(0.077)	(0.090)	
(log) Roman-road dist.				0.253***
				(0.069)
Pseudo R^2	0.116	0.126	0.134	0.118
Obs.	1855	1855	1855	1855

The coefficients are estimated using the specification of column 12 in Table 2.4. *Great brigandage* restricts our analysis to brigandage episodes occurring between 1860 and 1864. *Roman roads* considers distance computed along the Roman road network. Geographical, Distance and Municipal controls, Pop. growth and Province fixed effects are included in all specifications. Standard errors clustered at the district level in parentheses (* p<0.1, ** p<0.05, *** p<0.01). For a detailed description of the variables, see Appendix B.

and are reported consecutively (within the same document) in order to address potential over-reporting concerns. Second, we use alternative measures of municipal proximity. Specifically, we use linear distance without the log transformation; then, we use as our main independent variable the most discrete measure of distance, namely, the indicator for near-Piedmontese communities; and, finally, to show that our effect is not driven exclusively by the near-Piedmontese communities themselves, we implement our baseline specification by dropping these communities from our sample. Third, we present models alternative to the Negative Binomial regression. We repeat our analysis by estimating a Poisson model for our brigandage episodes and by using the zero-inflated versions of both Poisson and Negative Binomial regressions to account for the large number of zeros in our dependent variable. Additionally, we replace provincial indicators with more disaggregated indicators (at the district level) or more aggregated indicators (at the regional level). To allow for potential correlation in the error term, we cluster standard errors at the district and at the province levels, and we estimate a linear model taking into account spatial correlation. Finally, we show that our evidence is not driven by influential observations. We trim and winsorize the extreme 1% of observations of our dependent variable and consider alternative subsamples. The results of all these different specifications show that the association between brigandage intensity and proximity to near-Piedmontese communities is always statistically significant, negative and remarkably stable.

2.4.4 A cultural diffusion interpretation

Our results suggest a strong positive correlation between geographical distance from the closest near-Piedmontese community and the number of brigandage episodes that characterized the institutional rejection following the Italian unification. In this section, we discuss why historical interactions with neighboring communities made geographical distance a valid proxy of cultural vicinity to the Piedmontese and provide suggestive evidence that social interaction is indeed a relevant determinant of the phenomenon we observe.

Interactions with neighboring communities

Geographical isolation favored cultural persistence in near-Piedmontese enclaves but did not inhibit social interactions with neighboring populations. It is important for our empirical strategy that the descendants of Piedmontese settlers had intense contacts with local populations. There is an extensive literature documenting how culture and social norms spread through frequent contact and social interaction: in the language of Cavalli Sforza and Feldman (1981), both vertical and horizontal transmission mechanisms are assumed to co-exist.[35] For instance, Fogli and Veldkamp (2011) document the relationship between the increase in female labor force participation in the United States and the spacial diffusion of beliefs about women's roles. Spolaore and Wacziarg (2016*b*) claim that the historical fertility decline in Europe, which began in France, depended on the diffusion of new social norms and behavioral changes via social interaction and social influence. We conjecture that frequent interactions between near-Piedmontese enclaves and nearby communities generated a significant convergence in cultural and social norms and that the strength of those interactions strictly depended on geographic distance, considering in particular the lack of advanced communication systems in historical Southern Italy.

A first channel of social interaction was trade between the near-Piedmontese communities and the neighboring municipalities. Historical sources attest that some of the near-Piedmontese communities were not self-sufficient, necessitating frequent commercial exchanges in local markets and fairs. Bitonti (2012) documents how both Celle di San Vito and Faeto had frequent contacts with neighboring municipalities in the fifteenth century, particularly in establishing a commercial partnership with the town of Ariano Irpino. A second channel of interaction was intermarriage. Vegezzi Ruscalla (1862) reports that, after the Reformation and the following persecution, Guardia's citizens were sometimes forced to marry outside their own community. While there is no easily available record of

[35] For a review of the economic analysis of cultural transmission mechanisms, see Bisin and Verdier (2011).

local marriages, we may gain some insight by exploring the local diffusion of Gallo-Italic surnames. Rohlfs (1985) identifies 22 surnames that are likely to derive from historical Piedmontese migrations.[36] The author also reports the municipalities in which these surnames are relatively frequent. When we locate municipalities where these surnames are recorded, we observe that the average distance from the closest near-Piedmontese enclave is approximately 30 km, with a peak of 90 km. This evidence supports our conjecture that Piedmontese descendants had the opportunity to interact with nearby communities.

To further explore whether Piedmontese descendants had social relations with their neighbors, we again rely on linguistic similarities. In our specific environment, given the geographic isolation of our communities of interest, linguistic admixture is a distinct sign that some form of social interaction must have taken place. Mennonna (1987) states that the dialects of the near-Piedmontese enclaves of Basilicata had influences on the (non-Gallo-Romance) dialects of neighboring towns because of economic and cultural exchanges among bordering municipalities. Rohlfs (1988) reports Gallo-Romance influences in the – otherwise southern – dialects of the towns of Avigliano, Cancellara, Ruoti and Trivigno, close to the northernmost cluster of near-Piedmontese enclaves in Basilicata. Minor Gallo-Romance elements may also be found in the dialect spoken in Maratea, near the southernmost cluster (Figure 2.8).

The scope of cultural proximity

Table 2.6 presents the results of further analyses, clarifying the mechanisms behind our results. In columns 1–4, we estimate our main specification using municipalities within each quartile of the distribution of distances to the nearest near-Piedmontese community. Consistent with our idea that geographical distance facilitates social interaction and that the diffusion of cultural and social norms is non-linear (being much more intense at shorter distances), our estimates show that our global effect is driven primarily by municipalities in the first quartile of the distribution of distance. This finding means that the effect of being close to a near-Piedmontese community is strong only within a range of approximately 45 km from the latter. Note that, given the types of social interactions one may expect to have taken place in the Middle and Modern Ages in the areas under study (intermarriage, small-scale trade, participation in local markets and fairs), it seems plausible that the diffusion of near-Piedmontese cultural values from near-Piedmontese communities did not proceed beyond the limit of a day's walk (or donkey-ride) distance, which we may reasonably locate within a range of 40 to 50 kilometers.

[36] Figure 2.10 in Appendix A provides the original list of surnames by Rohlfs (1985).

If social interaction is indeed the mechanism through which geographic distance from a near-Piedmontese enclave impacts institutional rejection, we should observe that municipalities exposed to a higher number of Piedmontese descendants display a lower brigandage incidence. In columns 5 and 6, we test this hypothesis conditioning our sample to the presence of a near-Piedmontese enclave within a 50 km radius and exploiting two measures of intensity of exposure to near-Piedmontese population. The first variable (*Share of near-Piedmontese communities within 50 km*) measures for each municipality the number of near-Piedmontese communities over the total number of municipalities within a 50-kilometer radius neighborhood, weighted by inverse distance.[37] In addition to capturing the idea that cultural exposure increases in the frequency of near-Piedmontese communities located within a suitably large neighborhood, this variable incorporates the concept that as the closeness of near-Piedmontese enclaves increases, the chance of social interaction and, therefore, the cultural exposure increase. The second variable (*Share of near-Piedmontese population within 50 km*) also takes into account the size of municipalities and near-Piedmontese communities by weighting inverse distances by their population.[38] Again, the idea is that the likelihood of admixture with Piedmontese culture grows as the relative number of Piedmontese descendants with whom one might enter into contact increases. As one can see, the estimates of our coefficients retain a negative sign and statistical significance and are stable in size. For the sake of illustration, suppose that municipality i has 9 neighboring (that is, closer than 50 km) municipalities, all located 20 km from i and that none of those is a near-Piedmontese community. If a near-Piedmontese community were exogenously added to the set of neighbors at 20 km from the reference municipality, our estimate in column 5 predicts that the expected brigandage incidence in i would reduce by two-thirds. However, if, say, the added near-Piedmontese community were half as populated as the initial 9 neighbors, according to our estimate in column 6, the expected brigandage incidence in i would reduce by only two-fifths. We consider this finding to be a clear indication that cultural distance and the intensity of

[37] Let i be the reference municipality; N_i the set of municipalities that are closer than 50 kilometers to i; $d_{j,i}$ the distance between municipalities j and i; and NP the set of near-Piedmontese communities. Then

$$\text{Share of Near-Piedmontese communities within 50 km} = \frac{\sum_{j \in N_i \cap NP} d_{j,i}^{-1}}{\sum_{j \in N_i} d_{j,i}^{-1}}$$

[38] Allowing municipality j's population to be denoted by pop_j, we have

$$\text{Share of near-Piedmontese population within 50 km} = \frac{\sum_{j \in N_i \cap NP} d_{j,i}^{-1} \cdot \text{pop}_j}{\sum_{j \in N_i} d_{j,i}^{-1} \cdot \text{pop}_j}$$

the exposure to Piedmontese culture affected the strength of brigandage in the period in which brigandage was motivated primarily by institutional rejection, suggesting that higher exposure to Piedmontese culture induces a lower degree of institutional rejection.

Finally, one could expect the relative importance of social interaction with Piedmontese descendants to be less effective in larger cities. Indeed, as a municipality's population increases, the probability that a random inhabitant of the town will be subject to frequent contact with any specific cultural trait decreases.

Besides, the relative importance of interactions with Piedmontese descendants in determining the local cultural traits should be smaller, as exposure to different cultures is more frequent. Columns 7–9 include an indicator variable that equals one if the municipality is in the 75th, 90th and 95th percentiles of population, respectively, and its interaction with distance from the nearest near-Piedmontese community. Unsurprisingly, the coefficient on the interaction term is negative and decreases in magnitude as the size of the city increases, eroding the effect of exposure to Piedmontese culture.

2.5 Competing explanations

2.5.1 Alternative channels

The results presented thus far provide evidence that geographical proximity to a near-Piedmontese community was *ceteris paribus* associated with smaller brigandage incidence. The robustness of this result to changes in the model's specification and the addition of several geographical and socio-economic controls suggests that this finding was not due to observable peculiarities of near-Piedmontese communities. However, in this section, we test whether our results might be driven by pre-existing factors other than cultural proximity, such as differences in education, attitudes towards innovation and social capital.

Education. Higher levels of human capital may generate positive spillovers through social interactions and thereby influence the ease of adoption of new institutions around centers of diffusion. Our findings may hide the diffusion of superior education levels around near-Piedmontese communities. Overall, educational attainment in post-unitary Italy was very poor: according to the 1861 Census, more than 75% of the citizens were illiterate. This percentage rises to approximately 88% when we consider the continental territories of the former Kingdom of Two Sicilies. To control for potential differences in primary education and to rule out the potential correlation between education and the distance from the near-Piedmontese enclaves as a possible confounding factor, we collect

Table 2.6: Mechanisms

Dep. Variable: **Ep. p.m.**	By quartile of distance from nP community				Exposure intensity		City size		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(log) Dist. Piedm.	0.378***	0.192	-0.500	1.125			0.245**	0.236***	0.225***
	(0.076)	(0.522)	(1.044)	(2.310)			(0.110)	(0.084)	(0.085)
Share of nP comm. in 50 kms					-10.934***				
					(3.347)				
Share of nP pop. in 50 kms						-9.882**			
						(4.061)			
Cities (75percentile)							0.033		
							(0.300)		
Dist. Piedm. * Cities (75pctl)							-0.081		
							(0.074)		
Cities (90percentile)								0.683	
								(0.537)	
Dist. Piedm. * Cities (90pctl)								-0.236**	
								(0.118)	
Cities (95percentile)									0.971
									(0.634)
Dist. Piedm. * Cities (95pctl)									-0.298**
									(0.146)
Dist. quartile	1st	2nd	3rd	4th					
Pseudo R^2	0.174	0.0955	0.138	0.110	0.169	0.168	0.120	0.119	0.118
Obs.	464	464	464	463	516	516	1855	1855	1855

The coefficients are estimated using the specification of column 12 in Table 2.4. *Share of nP comm. in 50 kms* and *Share of nP pop. in 50 km* are described in Section 2.4.4. *Cities ("x"percentile)* is a dummy variable that equals 1 if the municipality has a population larger than the "x"th percentile. *Dist. Piedm. * Cities ("x"pctl)* is the interaction between *(log) Dist.Piedm.* and *Cities ("x"percentile)*. Geographical, Distance and Municipal controls, Pop. growth and Province fixed effects are included in all specifications. Standard errors clustered at district level in parentheses (* p<0.1, ** p<0.05, *** p<0.01).

information on the number of primary schools, pupils enrolled and teachers employed in primary schools in each municipality, exploiting a report by the Italian Statistical Office on the state of primary education during 1862-1863.[39] Column 1 in Table 2.7 reports our main specification including the number of schools reported in each municipality and the ratio between pupils and teachers as measures of, respectively, the diffusion and quality of primary education. Our coefficient of interest is not affected by the introduction of these controls.

Economic institutions. In this paragraph, we take into account two important dimensions of economic activity that may spread through space: agricultural productivity and financial institutions. First, since their establishment, near-Piedmontese communities and, through spatial diffusion, their neighbors might have specialized in agricultural activities that were less affected by the post-unification reforms. To account for this potential confounding factor, we take from the literature (Galor and Özak, 2016; Mayshar et al., 2016) and include controls for potential land outcomes and soil quality to capture differentials in land productivity.[40] In column 2 of Table 2.7, we show that our result is not dramatically affected by the addition of municipal-level controls referring to crop types, agricultural yields and raggedness, although the latter has some explanatory power in relation to our outcome. Second, the historical presence of a (relatively) developed financial sector may help cope with the short-run economic consequences of the institutional transplantation. Pascali (2016) links the development of modern banks to the presence of Jewish communities in the fifteenth century. In the same spirit, we inspect the impact of the presence of Jewish pawnshops in the south of Italy during the fifteenth century, before the ban imposed by the Spanish Crown, as documented by Montanari (1999). Column 3 of Table 2.7 shows that the results of the baseline specification are not affected by the inclusion of an indicator of the presence of a Jewish pawnshop in the municipality.[41]

Attitudes towards innovation. Disposition toward institutional innovation may differ across municipalities and territories. Southern Italy's history has been characterized by the succession of different rulers and crowns throughout the centuries, and there is evi-

[39]Educational reform in post-unification Italy was ineffective until Coppino's law in 1877. Primary schooling was offered at the municipal level until 1911. For further details, see Scarangello (1964).

[40]Our main soil quality measure is the caloric suitability index constructed by Galor and Özak (2016). We also make use of data from the European Soil Database (v. 2.0) compiled by the European Soil Data Center (ESDAC).

[41]The presence of Jewish communities may also capture potential differences in pre-existing attitudes towards individuals of different cultures and in relationships with the Church.

dence of many institutional changes that were opposed or accepted by local communities. Pre-existing attitudes associated with near-Piedmontese communities can be relevant in explaining our findings. In fact, one may suppose that reaction was weaker around near-Piedmontese areas not because of cultural similarities with their ancestors, as we claim, but because these communities were innately less inclined to rebel against institutional reforms. Corona (1995) reports instances of different behaviors by local communities in response to innovations in land property under the Bourbon regime. These reports allow us to construct a (censored) index of attitudes toward such innovations. Specifically, we assign a score of -1 to communities that enacted behaviors explicitly opposed to innovations and a score of $+1$ to municipalities displaying episodes of acceptance of innovative legislation (and a score of 0 to those for which there is a lack of evidence in either direction). Column 4 in Table 2.7 adds our attitude index to the baseline specification and shows that including this measure of past attitudes towards institutional change does not affect our result.

Social capital: charity lending. Another potential explanation for decreased brigandage intensity in near-Piedmontese communities is the possibility that these had developed internal social structures that led them to display a lower propensity to violently reject the institutional transplant. To assess the plausibility of this channel, we repeat our analysis controlling for the presence of charitable institutions in the community. We exploit the presence of the so-called *monti di pietà* (pawnbrokers) in the 16th and 17th centuries. These institutions were formal pawnbrokers run as charities: they were created with the explicit intention to aid the less fortunate. We collect these data from Montanari (1999) and create a dummy variable that equals one if a Monte di Pietà was reported in the municipality. In column 5 of Table 2.7, we report the results of the baseline specification including these controls. Our findings are not significantly affected by the new control variables.

Finally, in column 6 of Table 2.7 we control for all these variables at once and the coefficient of main interest is still positive and statistically significant.

2.5.2 Placebo tests

The results in the previous section show that controlling for available indicators of education, social capital, attitudes towards institutional change and economic and financial institutions at the municipal level does not alter our findings Thus, because formal institutions were homogeneous across Southern Italy at the time of the transplant and given

Table 2.7: Alternative explanations

	Additional controls for:					
Dep. Variable:	Education	Land quality	Fin. inst.	Attitude	Soc. capital	All
Ep. p.m.	(1)	(2)	(3)	(4)	(5)	(6)
(log) Dist. Piedm.	0.233***	0.179**	0.228***	0.226***	0.226***	0.191**
	(0.082)	(0.091)	(0.087)	(0.087)	(0.087)	(0.089)
Schools p.m.	0.164*					0.155*
	(0.084)					(0.085)
Pupils/teachers	0.003***					0.003***
	(0.001)					(0.001)
Raggedness		0.046***				0.044***
		(0.011)				(0.011)
Soil quality		-0.548				-0.504
		(0.667)				(0.647)
Agr. limitations		0.045				0.130
		(0.156)				(0.150)
Steepness		0.005				-0.058
		(0.082)				(0.082)
(log) Dist. Jewish			-0.155*			-0.146*
lenders			(0.083)			(0.081)
Prop. innovation				-0.081		-0.076
				(0.076)		(0.074)
Pawnshops					-0.074	-0.081
					(0.100)	(0.103)
Pseudo R^2	0.122	0.123	0.119	0.118	0.118	0.129
Obs.	1703	1826	1855	1855	1855	1674

The coefficients are estimated using the specification of column (12) in Table 2.4. The header of each column reports the alternative explanation addressed in the corresponding specification. Geographical, Distance and Municipal controls, Pop. growth and Province fixed effects are included in all specifications. Standard errors clustered at district level in parentheses (* p<0.1, ** p<0.05, *** p<0.01). For a detailed description of the variables, see Appendix B.

the absence of historical accounts of informal institutions distinguishing near-Piedmontese communities from neighboring settlements, the attribution of our result to cultural differences between near-Piedmontese and other communities appears sound; however, it may still be insufficient to conclude that the driving feature of our results lay in the cultural proximity of these near-Piedmontese communities with respect to the donor environment and that nearby municipalities experienced less brigandage because of the diffusion of cul-

tural traits springing from near-Piedmontese communities. First of all, near-Piedmontese municipalities' very status as a linguistic minority – rather than their specific ancestry – might be the reason for reduced reaction to new institutions in these towns. Second, lower brigandage in neighboring communities could have been reinforced not by cultural diffusion *from* the near-Piedmontese towns but *by sheer imitation of their behavior*. In this section, we conduct two placebo exercises that use as main independent variable the distance of each municipality from non-near-Piedmontese ethno-linguistic enclaves and their distance from municipalities that did not experience any brigandage episodes, respectively. The former exercise aims to rule out the possibility that our findings are driven by characteristics of cultural enclaves (e.g., social cohesion) that are nonetheless not necessarily exclusive of their Piedmontese heritage. The latter excludes the possibility that near-Piedmontese communities were (willingly or unwillingly) coordinating nearby municipalities into abstaining from rebellion and that they decided not to rebel by imitation.

Placebo: social capital and social cohesion. In a recent paper, Bracco et al. (2015) show that Albanian linguistic enclaves in Southern Italy display larger degrees of civicness than the average Southern municipality. This finding is attributed to the more intense *social cohesion* that emerges within ethno-linguistic enclaves, where bonds and relationships are formed by virtue of the enclave's cultural dissimilarity with respect to the surrounding environment. One might then suspect that a similar mechanism underlies our results, i.e., that near-Piedmontese communities had developed internal social structures that led them to display a lower propensity to violently reject the institutional transplant than their non-near-Piedmontese neighbors. To assess the plausibility of this channel, we repeat our analysis using non-near-Piedmontese linguistic enclaves as reference points. At present – as in the past – Southern Italy hosts 79 such enclaves, the majority of which (54) are of Albanian descent, with the remaining 25 being of Greek (23) or Croatian (2) origin. We inspect whether our results are replicated if distances from these enclaves are used.

Column 1 of Table 2.8 reports the result of the estimation of our main specification when the distance to the nearest non-near-Piedmontese ethno-linguistic enclave is used as the main independent variable. In column 2, we restrict the analysis only to those municipalities in the first quartile of the distribution of distance to the nearest non-near-Piedmontese community (we showed in Section 2.4.4 that our result is driven by municipality within a 45 km range from near-Piedmontese enclaves). Since Greek communities have a very particular geographic location (see Figure 2.1a) and are considerably distant from the

Table 2.8: Placebo regressions

Dep. Variable: **Ep. p.m.**	Non-Piedmontese		Albanian		Greek	
	(1)	(2)	(3)	(4)	(5)	(6)
(log) Dist. non-Piedm.	0.012	0.049				
	(0.068)	(0.054)				
(log) Dist. Albanian			0.068	0.021		
			(0.087)	(0.071)		
(log) Dist. Greek					-0.012	0.020
					(0.127)	(0.083)
Dist. quartile	All	1st	All	1st	All	1st
Pseudo R^2	0.117	0.167	0.117	0.179	0.117	0.110
Obs.	1855	464	1855	464	1855	464

Columns 1 and 2 consider all non-Piedmontese linguistic enclaves as reference points; columns 3 and 4 consider Albanian enclaves only; columns 5 and 6 consider Greek enclaves only. The even columns restrict the analysis to municipalities within the first quartile of distance from the nearest reference point. The dependent variable is Ep. p.m. The coefficients are estimated using the specification of column (12) in Table 2.4. Geographical, Distance and Municipal controls, Pop. growth and Province fixed effects are included in all specifications. Standard errors clustered at the district level in parentheses (* p<0.1, ** p<0.05, *** p<0.01). For a detailed description of the variables, see Appendix B.

areas where the near-Piedmontese communities lie, we tak into account the evidence on Albanian enclaves by Bracco et al. (2015) and repeat the analysis considering the two groups separately. All these specifications yield non-significant estimates for the coefficient of interest.[42] Thus, our attenuation effect of the proximity to a linguistic enclave on the intensity of brigandage appears only when the enclaves with respect to which such proximity is evaluated are the near-Piedmontese communities: this excludes the possibility that the channel of social cohesion due to the status of being an ethno-linguistic enclave drives our results.

Placebo: coordination and imitation. Even if one accepts the hypothesis that cultural proximity to the donor is the driving force behind less intense brigandage in near-Piedmontese communities, one may attribute the extension of such attenuation to

[42] As mentioned in Section 2.2.2, some Albanian and Greek communities were already recognized as linguistic enclaves in the 1861 Census. Our results are virtually the same when using the smaller sample of enclaves reported there (the Census recognizes 17 Albanian enclaves, instead of the 54 we use, and 9 Greek enclaves instead of 23). Results are available upon request.

neighboring communities to behavioral imitation of the near-Piedmontese rather than to 'cultural contagion' by them. If this were the case, we would expect the same phenomenon to arise in the vicinity of many areas with low brigandage intensity. In order to assess this, we repeat our analysis using as the main independent variable the distance from 200 randomly selected groups of 10 municipalities displaying no brigandage at all and located within 80 kilometers of a near-Piedmontese enclave (this step enables us to exclude using municipalities from macro-areas that were substantially unaffected by brigandage). If imitation of non-rebellious neighbors were a widespread phenomenon in the context under study, we would expect most of the estimates of our coefficient of interest to become positive. From Figure 2.3, we see instead that approximately 1.5% of our coefficients are significantly positive.[43] This finding suggests that, if it existed at all, the incentive to imitate the behavior of neighboring communities was not strong enough to significantly impact the intensity of resistance to the institutional transplant.

2.6 Brigandage and political participation

In this section, we examine the impact of the institutional rejection on political participation in the aftermath of the Italian unification. The imposition of the new institutions, the following unrest in response to the political regime change and the subsequent hard-handed repression by the national army are natural vehicles for the creation of a culture of mistrust and the reinforcement of negative attitudes toward the new political structure. Such an environment can lead to a decrease in cooperation among individuals and a reduction of trust in national institutions, resulting in lower political participation. Indeed, as shown by Lowes et al. (2017), a poor institutional environment can induce 'worse' cultural traits, which may in turn lead to socio-economic backwardness. Reminiscent of Banfield (1958) and Putnam (1993), we exploit the differences in civic capital across Italian territories and test whether a stronger institutional rejection, quantified by the intensity of brigandage, is associated with a weakening in the levels of political participation. In this respect, we follow the existing literature and proxy political participation by electoral turnout, the only measure available since 1861 at the local level.[44] In section 2.6.1, we study the relation between institutional rejection and turnout in the aftermath of the Italian unification, and in Section 2.6.2, we investigate whether effects of brigandage may

[43]Consider that while the near-Piedmontese communities themselves are excluded from the random reference sets of municipalities, positive coefficients should and will arise if most municipalities in the random set happen to be close enough to the near-Piedmontese enclaves.

[44]For a thorough review of the role of civic capital for economic development, see Guiso et al. (2011).

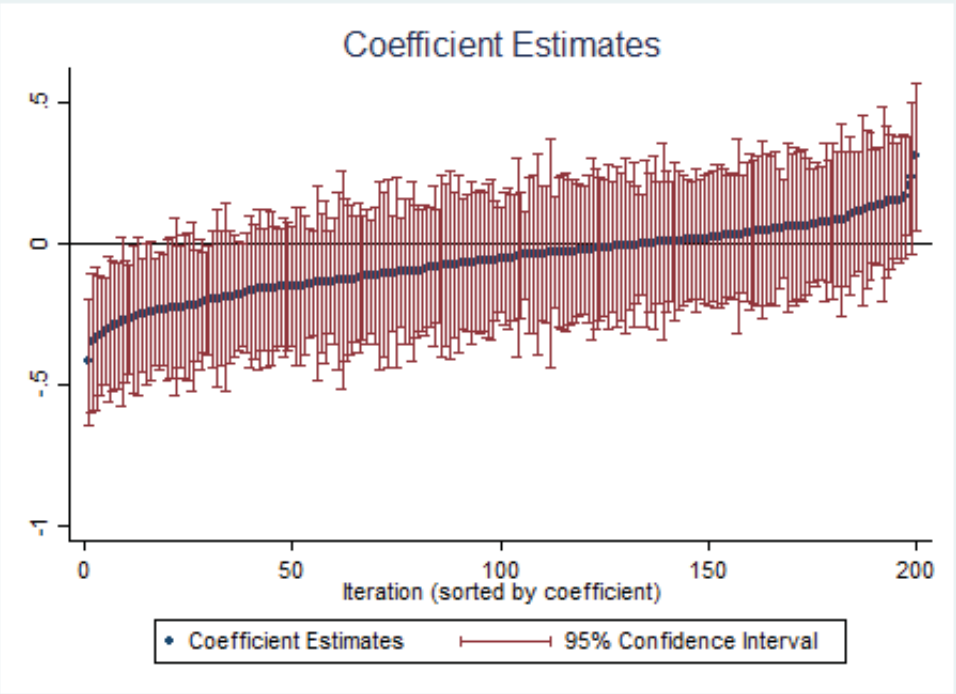


Figure 2.3: Coefficient estimates and 95% confidence intervals for 200 replications of the estimation exercise in column 12 of Table 2.4, where the main independent variable is the distance from a set of 10 randomly selected municipalities that (i) lie within 80 km of a near-Piedmontese community and (ii) have experienced no brigandage. Only 2 coefficients are positive and significant at 5% (3 at 10%).

be observed in the long run.

2.6.1 Immediate consequences of brigandage

A potential concern with this analysis is that a lack of trust in the transplanted institutions may have caused brigandage in the first place. In fact, lower turnout in areas characterized by more intense brigand activity may simply reflect their propensity to reject the transplanted institutions in one way (violent guerrilla conflict) or another (non-participation in elections). It is also possible that direct differential exposure to the military invasion generated distrust in the new regime, causing both lower levels of political participation and episodes of social unrest to cluster around the same areas. To address these and other related concerns, we show that episodes recorded in the years between the two earliest general elections following the Italian unification (held in 1861 and 1865) are associated with a relative reduction in electoral turnout form the first to the second rounds. As documented by our previous analyses, the patterns of local varia-

tions in brigandage were already clear in the early phases (1860-1861) of the reactionary movement. We thus assume, in this context, that turnout in the 1861 electoral round already incorporates the effects of local variations in the propensity for institutional rejection. By taking into account the relationship between differences in turnout between 1861 and 1865 and brigandage episodes occurring between 1862 and 1864, we are thus able to isolate the direct impact of the latter on political participation in addition to the pre-existing local differences in civic capital and trust in the new institutions.

In the analysis that follows, we exploit an additional data source: the *Atlante Storico Elettorale d'Italia*, a dataset collected by the Istituto Carlo Cattaneo containing detailed information on Italian national elections since 1861. During the period under analysis, municipalities were grouped into roughly equally sized constituencies (in terms of voters); each constituency elected one member of Parliament. Electoral data on turnout in elections held between 1861 and 1919 are available at the constituency level. All municipalities in which there was at least one eligible voter were reported. We use the Dizionario dei Comuni del Regno d'Italia (1863) to identify the formal constituency to which each municipality belonged, so we have a precise assessment of both turnout and the number of brigandage episodes by constituency at the same level of aggregation.[45] We then estimate the following linear model:

$$\Delta T_{i,j} = \alpha + \beta_1 Y_i + \beta_2' \mathbf{X}_i + \beta_3' \mathbf{H}_i + \gamma_j + \varepsilon_i \tag{2.2}$$

where $\Delta T_{i,j}$ is the difference in turnout between 1865 and 1861 in constituency i and province j; Y_i is the total number of episodes (per thousand inhabitants) that occurred between 1862 and 1864 in constituency i; $\mathbf{X_i}$ and $\mathbf{H_i}$ are vectors of geographical and constituency controls, respectively, with the former being measured for the main municipality of the constituency, while the latter are aggregated from municipal controls (see Section 2.3); γ_j is a province indicator;[46] and, finally, ε_i is an error term satisfying the usual assumptions. We expect the estimate of β_1 to be significantly different from zero; specifically, we conjecture that a higher incidence of brigandage should be negatively related to changes in turnout. The estimated coefficients are reported in Table 2.9.

[45] One needs to keep in mind that our outcome variable reflects only the political participation of the elites, who were less involved than popular masses in the rebellion against the Piedmontese. Indeed, according to the 1859 electoral law, eligible voters had to meet requirements in terms of age (older than 25), minimum literacy and income (at least 40 *lire* per year). The combination of these criteria reduced the number of potential voters in the country to 400,000, corresponding to approximately 2% of total population.

[46] Province fixed effects are particularly relevant in our specification since they capture the correlation between the Pica law (implemented at the provincial level) and electoral turnout, as underlined by Accetturo et al. (2017).

In columns 1–3, we regress the change in turnout on the brigandage episodes reported in the period between the two elections (1862-1864).

Table 2.9: Electoral turnout 1861-1865

Dep. Variable:	ΔT (1)	ΔT (2)	T_{1865} (3)	ΔT (4)	ΔT (5)	ΔT (6)	ΔT (7)
Ep. p.m. 1862-64	-0.025*	-0.039**	-0.035**	-0.046**	-0.041**		
	(0.011)	(0.015)	(0.013)	(0.017)	(0.016)		
Turnout 1861			0.168				
			(0.118)				
NP in constit.				-0.067			
				(0.074)			
Min. Dist. Piedm.					0.016		
					(0.026)		
Ep. p.m. 1860-61						-0.028	
						(0.054)	
Ep. p.m. 1865-70							-0.007
							(0.025)
Geog. Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Distance Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Municipal controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0709	0.126	0.297	0.122	0.121	0.0924	0.0876
Obs.	126	126	126	126	126	126	126

All coefficients are estimated by ordinary least squares. Dependent variables in column headings. *Ep. p.m. 1862-1864* measures episodes per 1,000 inhabitants occurring at the constituency level in infra-electoral years. T_{1865} and T_{1861} indicate turnout levels measured at the constituency level in 1865 and 1861 respectively. ΔT is the difference in turnout between the two years. Pop. 1861, Pop. growth and Province fixed effects are included in all specifications. Standard errors are clustered at the province level. (* p<0.1, ** p<0.05, *** p<0.01).

In column 1, we control for province fixed effects only, while in column 2, we add the geographical and municipal controls. In column 3, we use an alternative dependent variable, electoral turnout in 1865, but we include turnout in 1861 among the control variables. In columns 4 and 5, we include, respectively, a dummy variable that equals one if there is at least one near-Piedmontese enclave in the constituency and the (minimum) distance of the constituency from the nearest near-Piedmontese enclaves: together with the usual geographical and socio-economic controls, this corroborates the view that changes in turnout are affected from brigandage itself, rather than from its determinants. Finally, in columns 6 and 7, we show that using episodes not belonging to the infra-electoral period yields

non-significant estimates, reducing the concern for spurious correlation by highlighting the direct impact of episodes that may have effectively affected changes in individual behavior between 1861 and 1865. According to our estimate in column 2, one additional episode per thousand inhabitants is associated with a decrease of 3.9 percentage points in turnout between the two elections on average. This effect is sizable if we consider that the average change in turnout (in absolute value) at the constituency level was 11%, supporting the assumption that brigandage had a significant short-run effect on electoral participation.

2.6.2 The effects of brigandage in the long run

In the following section, we provide suggestive evidence that a negative relationship between the intensity of brigandage and electoral turnout persisted until the turn of the twentieth century. We repeat the exercise of column 3 in Table 2.9 using as the dependent variable turnout at each electoral round between 1892 and 1909. We restrict our

Table 2.10: Electoral turnout 1892-1909: Long Term Effects

Dep. Variable: **Turnout**	1892 (1)	1895 (2)	1897 (3)	1900 (4)	1904 (5)	1909 (6)
Ep. p.m.	-0.011*	-0.015*	-0.016**	-0.004	-0.009	-0.011
	(0.005)	(0.007)	(0.008)	(0.006)	(0.007)	(0.012)
Geog. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Distance Controls	Yes	Yes	Yes	Yes	Yes	Yes
Municipal controls	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.294	0.209	0.266	0.171	0.286	0.240
Obs.	119	119	119	119	119	119

All coefficients are estimated by ordinary least squares. All *Turnout* variables are measured at the constituency level. Turnout 1861, Pop. 1861, Pop. growth and Province fixed effects are included in all specifications. Geographical controls include (log) Area 1861 (ha), Altitude (100ms), Latitude and Longitude. Distance controls include (log) Dist. Naples, (log) Dist. n. prov. seat, (log) Dist. town, (log) Dist. coast, and (log) Dist. Papal States. Municipal controls include dummy variables for the presence in each municipality of civil, criminal or commercial courts; of the local episcopal or archiepiscopal seat; of secondary education institutes; of hospitals; and of relevant manufactures or proto-industrial plants. Standard errors are clustered at the province level. (* p<0.1, ** p<0.05, *** p<0.01).

attention to this period (thus ignoring elections held between 1865 and 1882) because it

provides us with rather homogeneous electoral rules[47] and a substantial expansion of the electorate in comparison to previous elections, increasing the validity of our results.[48] We also restrict our analysis to constituency that existed in both 1861 and the 1892-1909 period. In order to avoid using potential channels through which brigandage may have influenced turnout as regressors, we do not include measures after 1861 as controls. Table 2.10 reports estimates of β_1 when all episodes of brigandage are used as the main independent variable. The estimated coefficient is always negative, fairly stable and statistically significant until the turn of the century. These results indicate that an instance of short-run institutional rejection may have a long-lasting impact – up to half a century after its occurrence – on political participation, as it undermines trust in the institutions of the newborn state.

2.7 Conclusions

This work analyzes the outcomes following the institutional transplantation that took place in 1860-1865 in the context of the Italian unification process. The imposition of the Piedmontese administrative and legal framework on Southern Italy sparked an armed reaction by the peasant classes in continental Southern Italy, which is known as *brigandage*. We use historical sources and construct a dataset that allows us to evaluate the intensity of this phenomenon at the municipal level. Next, we relate this measure to the distance of each municipality from the nearest of ten communities descending from near-Piedmontese settlers. Across several model specifications and using a variety of controls, we find robust evidence that distance from the nearest near-Piedmontese community (which we use as a proxy for cultural distance) is *ceteris paribus* associated with a lower intensity and incidence of brigandage. We interpret this result as evidence in favor of the hypothesis that in the context of an institutional transplantation, cultural proximity with the donor environment reduces the propensity of the recipient environment for institutional rejection in the short term. Short-run rejection may have a direct impact on important social outcomes: here, we document a significant decline in electoral turnout in the aftermath of unification in places where more brigandage episodes were reported. Further suggestive evidence outlines a negative relationship between the intensity of institutional rejection and turnout that does not disappear until the turn of the twentieth century. This 40-year-long effect

[47]For instance, the three electoral rounds of 1882, 1886 and 1890 were organized with fewer than 40 constituencies, which makes a comparison with our original constituencies meaningless.

[48]Since the requirements for age, literacy and income were weakened, the new electoral rule more than doubled the number of people with the right to vote.

of brigandage suggests that the short-term rejection of Piedmontese institutions impacted the nation building process through its long-lasting influence on political participation. Overall, our results shed some light on the intricate relationship among cultural distance, institutional innovation and institutional effectiveness. Additionally, our analysis helps deepen the understanding of current political phenomena, such as electoral abstentionism, in light of the complex social dynamics triggered by the formation of contemporary European states.

2.8 Appendices

2.8.1 Appendix A

Additional figures and tables

Figure 2.4: Administrative map of continental Southern Italy following the Italian unification (1881). Source: `http://www.asciatopo.altervista.org/index.html`.

Figure 2.5: The Ancient Roman road network from McCormick et al. (2013) used to compute road distances. Municipalities which do not lie on any roads are projected onto the nearest point belonging to the network. Near-Piedmontese communities are denoted by the red pinpoints.

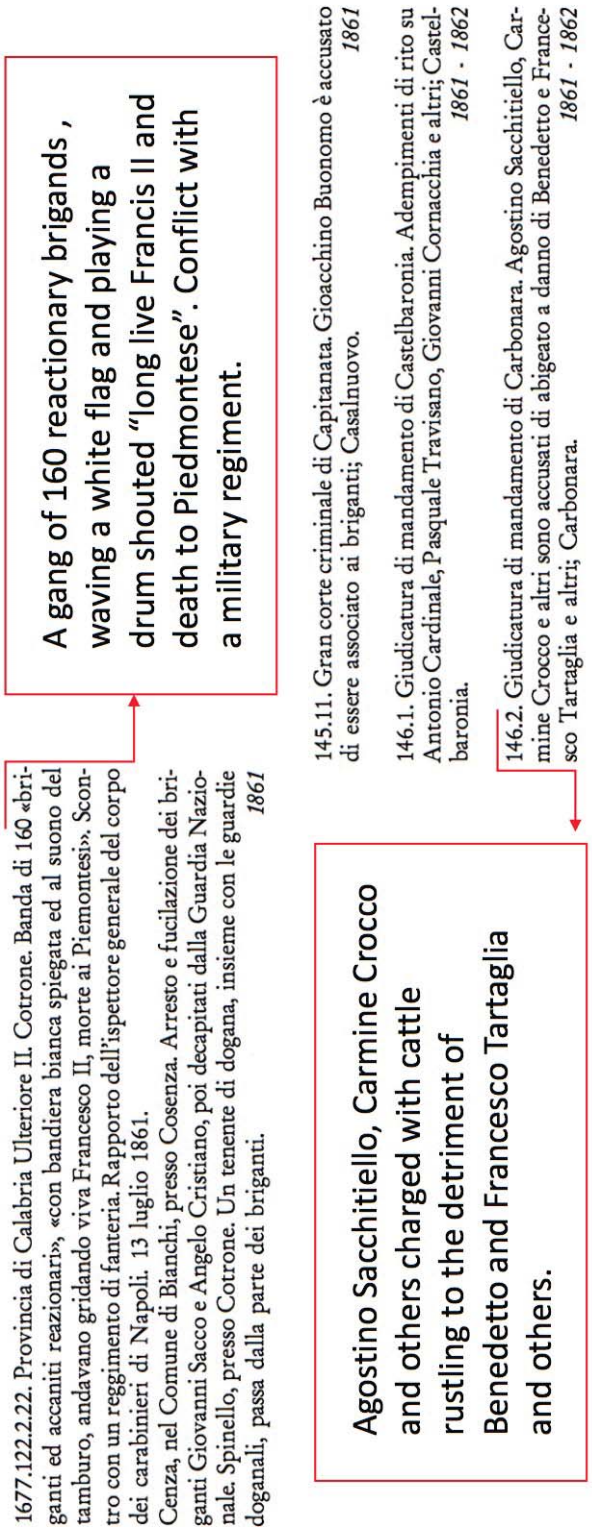


Figure 2.6: Excerpts from the State Archives used to construct our measure of brigandage intensity (Ministero per i Beni e le Attività Culturali, 1999-2001).

	Basilicata		Liguria		Piemonte	
	S. Chirico	Tito	Sassello	Borgomaro	Castelnuovo	Villafalletto
corpo	kuérpu	kórp	kóypu	kórpu	kórp	kórp
osso	ɥéssu	ɔssə	ósu	ósu	ós	ós
grosso	γrɥéssu	γróssu	grósu	grósu	grós	grós
grossa	gróssa	γróssa	grósa	grósa	grósa	grósa
coscia	kóssa	kúoša	kóeša	kóeša	kóeša	kóesa
macina	máćina	[súola]	móera	móeya	móra	—
cuore	kórᵘ	kúorə	kóe	kóe	kóer	koér
nuora	nóra	núora	nœérya	nœéya	nóla	nóra
sorella	sórə	súora	sœé	sœé	[suréla]	sœéri
fuori	fóra	fúora	fóera	defóera	fóra	fóra
piove	ˈćóvitᵘ	ˈćúovə	ćóeve	ćóeve	pyœéve	pyœévi
fuoco	fɥéku	fúoγu	fœ	fóegu	fœgu	fóe
oggi	óyi	úoyi	aŋkóe	aŋkœéy	aŋkœ	ŋkóe

Figure 2.7: Examples of linguistic similarities between Lucanian, the Gallo-Italic dialect of Tito, Western Ligurian and Southern Piedmontese. Source: Rohlfs (1988).

Figure 2.8: Maps of the Gallo-Romance clusters in Basilicata. Source: Rohlfs (1972).

Italiano	*Dialetto di Guardia*	*d'Angrogna*	*di Cosenza*
Mio padre	Mon paire	Mon paire	Patrima
Mia madre	Ma maire	Ma maire	Mammama
Mio avo	Mon Donn	Mon Donn	Nannuma
Mia ava	Ma gnogna	Ma nonna	Nannama
Tuo zio	Ton barba	Ton barba	Ziuma
Tua zia	Ta dant	Ta dant	Ziama
Suo fratello	Son fraire	Son fraire	Frate suo
Sua sorella	Sa sorr	Sa sorr	Suoru sua
Tuo nipote	Ton nibù	Ton nebu	Neputìta
Tua nipote	Ta nessa	Ta nessa	Id.
Mio suocero	Mon sière	Mon messer	Suocruma
Mia suocera	Ma madona	Ma madona	Socrama

Figure 2.9: Examples of linguistic similarities between Guardia Piemontese and Val d'Angrogna (Piedmont). Source: Vegezzi Ruscalla (1862).

53. L'elemento galloitalico. Alle immigrazioni piemontesi (v. num. 4) si possono attribuire i seguenti cognomi: *Antonaglia, Balasco, Bargellino, Braida, Cerone, Clausi, Di Lema, Doti, Mongiovi, Orenga, Piccitto, Posca, Ribba, Sadotti, Saluzzi, Sanchìrico, Tomasco, Vainieri, Varallo, Varanzano, Vergalliti, Verna.*

Figure 2.10: List of surnames attributed to the Piedmontese migration in Basilicata. Source: Rohlfs (1985)

Table 2.11: Table 2.4, Columns (1)-(6), reporting all coefficients

Dep. Variable:	Total Number of Episodes (Ep. tot)					
	(1)	(2)	(3)	(4)	(5)	(6)
(log) Dist. Piedm.	1.796**	1.752**	1.517**	0.185***	0.178***	0.169***
	(0.774)	(0.773)	(0.720)	(0.057)	(0.058)	(0.058)
(log) Pop. 1861	5.150***	5.094***	3.245***	0.726***	0.714***	0.650***
	(0.665)	(0.662)	(0.476)	(0.060)	(0.060)	(0.059)
Pop. growth 1824-61	1.496***	1.498***	0.381	0.021**	0.022**	-0.026
	(0.167)	(0.167)	(0.467)	(0.009)	(0.009)	(0.019)
Latitude	-1.727	-1.319	-0.562	-0.667***	-0.578***	-0.536***
	(1.550)	(1.580)	(1.579)	(0.155)	(0.156)	(0.148)
Longitude	0.200	-0.066	-0.345	0.133	0.069	0.082
	(1.014)	(1.028)	(0.966)	(0.132)	(0.132)	(0.131)
(log) Area 1861	3.217***	3.293***	2.642***	0.541***	0.556***	0.538***
	(0.361)	(0.365)	(0.314)	(0.052)	(0.052)	(0.051)
Altitude (100m)	0.513***	0.472***	0.445***	0.083***	0.072***	0.071***
	(0.154)	(0.155)	(0.149)	(0.018)	(0.018)	(0.018)
(log) Dist. Naples		-1.268	-1.796**		-0.295**	-0.393***
		(1.237)	(0.760)		(0.120)	(0.114)
(log) Dist. seat		-1.048*	0.111		-0.014	0.081
		(0.542)	(0.409)		(0.050)	(0.049)
(log) Dist. town		-0.998*	-0.449		-0.049	-0.020
		(0.584)	(0.556)		(0.043)	(0.042)
(log) Dist. coast		0.947***	1.154***		0.131***	0.153***
		(0.297)	(0.262)		(0.036)	(0.035)
(log) Dist. Papal States		-0.315	0.016		-0.122	-0.115
		(0.670)	(0.602)		(0.080)	(0.078)
Archbishop			8.071*			0.580***
			(4.701)			(0.174)
Bishop			8.582**			0.056
			(3.454)			(0.103)
Civil court			-4.286			0.091
			(11.933)			(0.477)
Criminal court			31.461***			1.391***
			(11.976)			(0.417)
Comm. court			4.144			-2.438***
			(13.267)			(0.714)
Liceo			10.441			-0.038
			(9.847)			(0.333)
Collegio			-0.710			-0.314
			(11.116)			(0.380)
Sec. school			-1.788			0.011
			(2.486)			(0.136)
Manufacture			2.054			-0.093
			(2.964)			(0.152)
Hospital			6.551			0.193
			(7.463)			(0.159)
Model	OLS	OLS	OLS	Neg. Bin.	Neg. Bin.	Neg. Bin.
Provincial controls	Yes	No	No	Yes	No	No
Region FEs	Yes	No	No	Yes	No	No
Province FEs	No	Yes	Yes	No	Yes	Yes
Adj./Pseudo R^2	0.422	0.423	0.488	0.155	0.157	0.163
Obs.	1855	1855	1855	1855	1855	1855

Table 2.12: Table 2.4, Columns (7)-(12), reporting all coefficients

Dep. Variable:	Episodes per thousand inhabitants (Ep. p.m.)					
	(1)	(2)	(3)	(4)	(5)	(6)
(log) Dist. Piedm.	0.717***	0.706***	0.699***	0.225***	0.222***	0.227***
	(0.164)	(0.164)	(0.165)	(0.058)	(0.059)	(0.059)
Pop. growth 1824-61	0.006	0.005	-0.050	0.006	0.006	-0.050**
	(0.051)	(0.052)	(0.072)	(0.009)	(0.009)	(0.020)
Latitude	-0.921**	-0.799*	-0.823*	-0.657***	-0.574***	-0.550***
	(0.430)	(0.436)	(0.438)	(0.162)	(0.164)	(0.159)
Longitude	-0.292	-0.369	-0.362	0.119	0.057	0.070
	(0.245)	(0.245)	(0.248)	(0.136)	(0.136)	(0.136)
(log) Area 1861	0.593***	0.607***	0.584***	0.403***	0.411***	0.387***
	(0.075)	(0.075)	(0.080)	(0.041)	(0.040)	(0.043)
Altitude (100m)	0.186***	0.174***	0.175***	0.094***	0.085***	0.084***
	(0.047)	(0.048)	(0.048)	(0.018)	(0.018)	(0.018)
(log) Dist. Naples		-0.445**	-0.474***		-0.288**	-0.359***
		(0.181)	(0.180)		(0.119)	(0.116)
(log) Dist. seat		0.018	0.050		-0.013	0.061
		(0.074)	(0.079)		(0.050)	(0.052)
(log) Dist. town		0.222***	0.239***		0.042	0.068*
		(0.076)	(0.079)		(0.041)	(0.039)
(log) Dist. coast		0.199***	0.211***		0.140***	0.158***
		(0.063)	(0.064)		(0.037)	(0.036)
(log) Dist. Papal States		-0.055	-0.059		-0.121	-0.120
		(0.161)	(0.162)		(0.080)	(0.079)
Archbishop			0.996			0.453**
			(0.804)			(0.179)
Bishop			0.121			-0.080
			(0.318)			(0.105)
Civil court			-0.263			0.032
			(1.059)			(0.539)
Criminal court			1.480			1.469***
			(1.042)			(0.434)
Comm. court			-1.673			-2.577***
			(1.150)			(0.803)
Liceo			-0.738			-0.159
			(0.855)			(0.368)
Collegio			-0.415			-0.509
			(0.934)			(0.383)
Sec. school			-0.189			-0.067
			(0.266)			(0.138)
Manufacture			0.176			-0.243
			(0.475)			(0.152)
Hospital			-0.286			0.095
			(0.683)			(0.156)
Model	OLS	OLS	OLS	Neg. Bin.	Neg. Bin.	Neg. Bin.
Provincial controls	Yes	No	No	Yes	No	No
Region FEs	Yes	No	No	Yes	No	No
Province FEs	No	Yes	Yes	No	Yes	Yes
Adj./Pseudo R^2	0.352	0.354	0.353	0.111	0.113	0.118
Obs.	1855	1855	1855	1855	1855	1855

Robustness checks

This section presents a series of checks to verify the robustness of the baseline estimates. First, we use alternative proxies for the dependent variable and the main independent variable. Then, we test our model specification both in terms of model and controls. The estimates of the main coefficient of interest in these alternative specifications are reported in Table 2.13.

The first three specifications, the estimates of which we present in panel (a) of Table 2.13, refer to variations in the dependent variable, i.e., the intensity of brigandage. Row 1 uses episodes reported by courts and we find a positive but not significant effect. Similarly, a non-significant effect is the one obtained using data from other sources, which include ecclesiastical authorities, the Italian Ministry of Justice and other miscellaneous sources, including the military (row 2). In row 3 we repeat our analysis using versions of the dependent variable that we deem particularly significant. We recode our dependent variable aggregating all the entries that share the same date and the same place and are reported consecutively (within the same source), as these distinct reports might be originally generated by the same event.[49] The new definition of brigandage episodes reduces the number of episodes per municipality and addresses the potential concern of over-reporting. The coefficient is similar in magnitude and significance to that found in our previous results.

In panel (b) of Table 2.13, we propose three alternative measures of municipal proximity. In row 4, we use linear distance without taking the logarithm. As explained above, the use of the log-transformed version of this distance was *a priori* motivated by the fact that cultural similarity decreases more than linearly once one moves farther away from the center of cultural diffusion and by our desire to use geographical distance as a proxy for cultural distance. For the sake of robustness, we show that our result holds even when the simpler kilometric distance is used. Rows 5 and 6 show the impact of near-Piedmontese communities themselves on our result. In row 5, we use as our main independent variable the most discrete measure of distance, namely the indicator for near-Piedmontese communities. As one might expect, being a near-Piedmontese community has a strong negative effect on brigandage incidence. Nevertheless, our effect is not only driven by the centers of highest cultural proximity with the donor: in row 6, we use our main specification while dropping near-Piedmontese communities from our sample. While the absence of the units for which the effect is strongest reduces the magnitude of our coefficient, we nevertheless observe a sizable effect. This again confirms that cultural and social norms

[49] It could be the case that different convictions following a single guerrilla event are reported in multiple entries.

indeed spread from near-Piedmontese enclaves to Southern Italy municipalities and that the degree of cultural similarity may then have determined, *ceteris paribus*, the intensity of institutional rejection in the form of brigandage. This evidence thus also justifies ex post our choice of using a continuous rather than discrete measure of proximity.

Rows 7, 8 and 9 present model alternatives to the Negative Binomial regression of model 12 in Table 2.4. As Poisson regression is likely the most widely used count model, we repeated our analysis by estimating a Poisson model for our brigandage episodes. Moreover, we used the zero-inflated versions of both Poisson and Negative Binomial regressions to account for the large number of zeros in our dependent variable (just over one-third of Southern Italy municipalities experienced no brigandage). The following specifications refine our baseline model by adding further controls or by imposing less-stringent assumptions. Row 10 shows that our result holds when replacing provincial indicators with more disaggregated indicators (at the district level).[50] To allow for potential correlation in the error term, we consider clustered standard errors at the district or at the province level (rows 11-12). Finally, in row 13 we implement the linear model of column 9 in Table 2.4 taking into account spatial correlation.[51] The findings show that our result is significant and robust across the different specifications

Finally, in panel d), we show that our evidence is not driven by influential observations. First, we trim (row 14) and winsorize (row 15) the extreme 1% of observations of our dependent variable. Furthermore, in rows 16–18, we consider alternative subsamples. In row 16, we exclude cities with more than 10000 inhabitants and province seats. We then exclude all the municipalities close[52] to Naples and the Papal States' border to rule out the possibility that our results are driven by the areas that received support by the former king and the Church. These different specifications show that the distance from near-Piedmontese community is always statistically significant, negative and remarkably stable.

[50]Both in the pre-unification kingdom and in post-unification Italy, the 16 provinces of Southern Italy were the main administrative units. The main local government units were the municipalities, which constitute our reference population. The 57 districts were intermediate administrative subdivisions determined by the domains of jurisdiction of ordinary courts.

[51]In particular, we rely on the function used by Hsiang (2010).

[52]We consider all the municipalities in the top 5% of the distribution of distances to Naples and the Papal States.

Table 2.13: Robustness checks

	$\hat{\beta}_1$	s.e.	Obs.
a) Dependent Variable			
1) Source: Courts	0.172	(0.148)	1,855
2) Source: Other	0.438	(0.272)	1,855
3) Different aggregation	0.220***	(0.077)	1,855

The coefficient refers to the variable (log) Dist. Piedm. and is estimated using the specification of column (12) in Table 2.4; Geographical, Distance and Municipal controls, Pop. growth and Province FEs are included in all specifications; standard error clustered at district level in parentheses (* p<0.1, ** p<0.05, *** p<0.01). *Different aggregation* excludes a number of reports which may be suspected to refer to the same episodes being signaled by different authorities.

b) Proximity Variable			
4) Distance in kms	0.006**	(0.003)	1,855
5) Near-Piedmontese dummy	-0.926***	(0.234)	1,855
6) Excluding near-Piedmontese	0.200*	(0.109)	1,845

The dependent variable is Ep. p.m.; the coefficient refers to the variable indicated on the left and is estimated using the specification of column (12) in Table 2.4; Geographical, Distance and Municipal controls, Pop. growth and Province FEs are included in all specifications; standard error clustered at district level in parentheses (* p<0.1, ** p<0.05, *** p<0.01). *Distance in kms* uses linear distance in kilometers from the nearest nP community, instead of its logarithm. *Near-Piedmontese dummy* substitutes the distance measure with the indicator of whether the municipality is a nP community. *Escluding near-Piedmontese* replicates our baseline analysis excluding the 10 nP communities.

c) Specification			
7) Poisson	0.215***	(0.055)	1,855
8) Zero-inflated Neg. Bin.	0.227***	(0.086)	1,855
9) Zero-inflated Poisson	0.230***	(0.051)	1,855
10) District FEs (57) + robust s.e.	0.167**	(0.071)	1,855
11) Clustering at Province level	0.227***	(0.076)	1,855
12) Clustering at Region level	0.227***	(0.066)	1,855
13) Spatial correlation	0.699***	(0.163)	1,855

The dependent variable is Ep. p.m.; the coefficient refers to the variable (log) Dist. Piedm. and is estimated using the specification indicated on the left. Unless differently specified: Geographical, Distance and Municipal controls, Pop. growth and Province FEs are included in all specifications; standard error clustered at district level in parentheses unless differently specified. (* p<0.1, ** p<0.05, *** p<0.01). *Spatial correlation* applies Conley standard errors (For more detail on the function implemented see Hsiang, 2010). For a detailed description of the variables, see Appendix B.

d) Alternative samples			
14) Trimming	0.219***	(0.085)	1,837
15) Winsorizing	0.223***	(0.085)	1,855
16) No large towns	0.238***	(0.084)	1,765
17) No Naples neighborhood	0.198**	(0.088)	1,769
18) No Papal States neighborhood	0.183*	(0.101)	1,730

The dependent variable is Ep. p.m.; the coefficient refers to the variable (log) Dist. Piedm. and is estimated using the specification of column (12) in Table 2.4; Geographical, Distance and Municipal controls, Pop. growth and Province FEs are included in all specifications; standard error clustered at district level (* p<0.1, ** p<0.05, *** p<0.01). *Trimming* and *Winsorizing* refer to, respectively, truncation and censoring of our measure of brigandage intensity at the 99th percentile. In *No large towns* we drop all province seats and all towns with more than 10,000 inhabitants in 1861. *No Naples neighborhood* and *No Papal States neighborhood* exclude all municipalities lying within 25 kilometers from, respectively, Naples and the northern border.

2.8.2 Appendix B

Variables

This table summarizes the variables used in our analysis, with a brief description, sources and summary statistics.

Variable	*Definition*	*Source*	*Range*	*Mean*
Panel a. Dependent variables.				
Ep. (tot)	Number of reported episodes of brigandage at municipal level	Italian Ministry of Culture, State Archives	0 – 211	6.6
Ep. p.m.	Ep. tot per 1,000 inhabitants (Pop. 1861)	Italian Ministry of Culture, State Archives	0 – 64.13	1.93
Panel b. Population Data.				
Pop. 1861	Population reported for the municipality in 1861	Italian 1861 Census	129 – 471,463	3,840
Pop. pre-1861	Data collected during the Kingdom of Two Sicilies in 1850s and published by the Italian Statistical Office after the unification	Dizionario dei Comuni del Regno d'Italia, 1863	116 – 447,065	3,658
Pop. growth	Growth rate of Pop. 1861 relative to Pop. pre-1861		-0.66 – 22.86	-0.04
Panel c. Geographical Controls.				
Area 1861	Area of municipal territory (hectares)	Archivio climatico DBT, ENEA	10 – 59,335	4,164
Altitude	Altitude of the municipal seat (ms)	Archivio climatico DBT, ENEA	1 – 1,433	415
Lat/Lon	Coordinates of the municipal centroid	Archivio climatico DBT, ENEA		
Panel d. Distance Controls.				

Variable	*Definition*	*Source*	*Range*	*Mean*
Dist. Naples	Linear distance from Naples (kms)	Based on coordinates from Archivio climatico DBT, ENEA	0 – 367	160
Dist. n. prov. seat	Linear distance from the nearest province seat (kms)	Based on coordinates from Archivio climatico DBT, ENEA	0 – 108	32
Dist. town	Linear distance from the nearest town with at least 10,000 inhabitants (kms)	Based on coordinates from Archivio climatico DBT, ENEA	0 – 77	19
Dist. coast	Linear distance from the closest maritime municipality (kms)	Based on coordinates from Archivio climatico DBT, ENEA	0 – 97	22
Dist. Papal States	Linear distance from the border with the Papal States (kms)	Based on coordinates from Archivio climatico DBT, ENEA	0.2 – 452	192
Dist. Piedm.	Linear distance from the closest near-Piedmontese community	Based on coordinates from Archivio climatico DBT, ENEA	0 – 228	89
Panel d. Municipal Controls.				
Archbishop	A dummy variable that equals 1 if an archbishopric is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.013
Bishop	A dummy variable that equals 1 if a bishopric is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.043

Variable	*Definition*	*Source*	*Range*	*Mean*
Civil Court	A dummy variable that equals 1 if a "Gran Corte Civile" is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.022
Criminal Court	A dummy variable that equals 1 if a "Gran Corte Criminale" is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.081
Comm. Court	A dummy variable that equals 1 if a "Tribunale di Commercio" is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.002
Liceo	A dummy variable that equals 1 if a "Liceo" (high school) is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.003
Collegio	A dummy variable that equals 1 if a "Collegio Reale" (high school) is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.006
Sec. School	A dummy variable that equals 1 if a "Scuola Secondaria" (middle school) is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.022
Manufacture	A dummy variable that equals 1 if a "Fabbrica Notabile" (Relevant Manufacture) is reported in the municipality by 1839, 0 otherwise	Statistiche dell'Italia by L. Serristori, 1839	{0,1}	0.017
Hospital	A dummy variable that equals 1 if an hospital was established in the municipality between 1831-1856, 0 otherwise	Annali Civili of the Kingdom of Two Sicilies, 1857	{0,1}	0.015
Panel e. Province-level controls.				
Young men	Number of men aged 19-25 eligible for military draft in 1834-35	Annali Civili of the Kingdom of Two Sicilies, 1836	12,664 – 48,561	29,403

Variable	*Definition*	*Source*	*Range*	*Mean*
Real Estate	Number of "possidenti" (real estate and land owners) in 1834-35	Annali Civili of the Kingdom of Two Sicilies, 1836	22,690 – 124,944	70,275
Liberal Arts	Number of liberal arts professional workers in 1834-35	Annali Civili of the Kingdom of Two Sicilies,1836	1578 – 8755	4,283
Farmers	Number of farmers in 1834-35	Annali Civili of the Kingdom of Two Sicilies, 1836	69,931 – 233,346	128,807
Artists	Number of artisans and house servants in 1834-35	Annali Civili of the Kingdom of Two Sicilies, 1836	2,655 – 94,687	18,422
Fishermen	Number of fishermen and sailors in 1834-35	Annali Civili of the Kingdom of Two Sicilies, 1836	250 – 25,530	2,829
Duties	Average amount of money collected with the duties on milled grain between 1850 and 1851	Annali Civili of the Kingdom of Two Sicilies, 1852	8,569 – 130,291	52,277
Revenues	Average revenues between 1850 and 1851	Annali Civili of the Kingdom of Two Sicilies, 1852	110,397 – 839,527	350,960
Public Expenditure	Average municipal public expenditures between 1850 and 1851	Annali Civili of the Kingdom of Two Sicilies, 1852	11,136 – 193,766	60,695
Suppressed monasteries	Number of suppressed monasteries between 1806-1815 at the province level	Villani (1964)	30 – 202	101.24
Reinstated monasteries	Number of reinstalled monasteries in 1818 at the province level	Villani (1964)	4 – 66	25.70

Variable	*Definition*	*Source*	*Range*	*Mean*
Buyers per contract	Ratio between the number of buyers and the number of transactions concerning the alienation of national wealth between 1806 and 1815 at the province level	Villani (1964)	0.36 – 0.87	0.57
Rent per contract	Ratio between total rent (1,000s *lire*) and the number of transactions concerning the alienation of national wealth between 1806 and 1815 at the province level	Villani (1964)	0.060 – 0.584	0.13
Panel f. Other controls.				
Schools p.m.	Number of primary schools per 1,000 inhabitants in the municipality during the school year 1862-1863	Istruzione Elementare Pubblica per Comuni: Anno Acolastico 1862-1863, 1865	0 – 5.28	0.78
Pupils/teachers	Number of pupils per teacher in primary schools in the municipality during school year 1862-1863	Istruzione Elementare Pubblica per Comuni: Anno Acolastico 1862-1863, 1865	0 – 461	37.7
Raggedness	Maximum difference in altitude within municipal territory (ms)	Archivio climatico DBT, ENEA	7 - 2,534	726
Soil Quality	Caloric suitability index	Galor and Özak, 2016	8.01 – 8.61	8.28
Agr. limitations	A dummy variable that equals 1 if limitations to the agricultural use where detected in the municipality, 0 otherwise	European Soil Database (v.2.0) by the European Soil Data Center (ESDAC)	{0,1}	0.81
Steepness	Dominant slope of the territory, ranging between 1 (Level - 0-8%) to 4 (Steep- over 25%)	European Soil Database (v.2.0) by the European Soil Data Center (ESDAC)	1 – 4	2.69

Variable	*Definition*	*Source*	*Range*	*Mean*
Propensity to innovation	Behaviors by local communities in response to innovations in land property under the Bourbon regime in the last decades of the eighteenth century. This variable equals -1 if the municipality explicitly opposed to innovations, $+1$ if the municipalities cooperated with innovative legislation, 0 otherwise.	Corona (1995)	{-1, 0, 1}	0.04
Pawnshops	Presence of pawnshops (*monti di pietà*) and other charitable funds in the 16th and 17th centuries	Montanari (1999)	{0,1}	0.005
Jewish lenders	Presence of Jewish moneylenders in the 15th century	Montanari (1999)	{0,1}	0.016

Chapter 3

Corruption and personnel selection and allocation in the public sector

Co-authored with Sauro Mocetti.

3.1 Introduction

Corruption is widely believed to entail large economic and social costs. The economic literature has so far explored several channels through which corruption may affect economic outcomes. Some authors highlight its effects in terms of distortion of private decisions, such as investments (Shleifer and Vishny, 1993; Mauro, 1995) and human capital accumulation (Mo, 2001). Others focus on the activities of the public sector, documenting relationships between corruption and inefficiencies in the composition of government expenditure (Mauro, 1998), lower productivity of public investments (Del Monte and Papagni, 2001) and higher shares of goods and services procured by the public administration on non-competitive markets (Hessami, 2014).

In this paper we analyze the impact of corruption (as measured by local-level statistical indicators based on the number of reported crimes, on citizens' trust in local public institutions, on perception of administrations' integrity and on the quality of public expenditure) on personnel selection and allocation in the public sector. More specifically, we address two issues: first, we examine whether corruption affects the *selection into the public sector* of individuals with different levels of (observable) human capital; second, we examine the relationship between corruption and the *allocation within the public sector* of differently educated individuals to jobs with different skill content. Poorer recruitment

and misallocation of human resources within public agencies might have significant and long-lasting consequences on the quality of the administration's economic decisions and on the effectiveness of the services provided by the public sector. Nevertheless the impact of corruption through these channels was surprisingly uninvestigated.

The empirical analysis is based on two complementary data sources containing information on Italian public and private employees and exploits several measures of corruption. We examine whether areas characterized by higher values of our corruption indicators show peculiar patterns of skill-based selection into and allocation within the public sector. Although we use cross-sectional variation, our empirical strategy mirrors a difference-in-differences approach where the "treatment" is represented by the intensity of corruption at the local level and exposure to the treatment is determined by individual characteristics of the employees (e.g. whether they are employed in the public or in the private sector). Time-invariant heterogeneity that might be correlated to both corruption and human capital endowments is captured by the inclusion of fixed effects at the local level. Moreover, to address reverse causality – the possibility that corruption itself be the consequence of poor selection and allocation of human resources by the public administration – we instrument corruption with past dependence upon public spending and past domination spells, i.e. with factors likely to be correlated with corruption but predating the hiring of current public employees.

We find that public employees are, on average, more educated and obtained higher grades at school than their professional counterparts in the private sector. However, in areas with higher values of our corruption indicators the relationship between educational attainments and the likelihood of joining the public sector is substantially weaker. The negative impact of corruption is concentrated among those with higher skill content jobs, such as managers and highly skilled professionals. As for the allocation process, we find that a higher level of the corruption indicators is associated with an increase, relative to the private sector, in the likelihood of mismatch between individual educational attainments and the skill content of the job one is assigned to. This mismatch comes mostly in the form of under-education – individuals being assigned to jobs that are, on average, undertaken by more qualified personnel – rather than over-education. We also show that mismatch is not merely a "mechanical" consequence of poorer selection processes, nor of inflation in the number of managerial positions. Finally, we show that, beyond worse selection and allocation processes, corruption also leads to a relatively lower effort by public employees, as measured by hours worked and absenteeism.

The literature has already partially dealt with the relationship between corruption and occupational choices. Murphy et al. (1991) and Acemoglu and Verdier (1998) argue that corruption magnifies rewards to rent-seeking activities, thus subtracting valuable human resources to entrepreneurship and distorting the allocation of talent across sectors. Concerning selection into the public sector, experimental evidence suggests that more corrupt environments encourage entry by the dishonest into the public sector: Banerjee et al. (2015) and Hanna and Wang (2017) find negative self-selection into the Indian public administration, while Barfort et al. (2017) find positive self-selection into the Danish public sector.

Our paper innovates upon the existing literature along several directions. First, the economic impact of corruption has typically been investigated using cross-country evidence (at a single point in time). However, the cross-sectional relationship might be severely biased as corruption and the other variables of interest are likely to have common correlates that cannot all be credibly controlled for: stated differently, less corrupt societies appear to perform well in almost any dimension, and the risk of bias due to an omitted variable (e.g. of institutional or cultural nature) is large. To address this issue, some papers introduce country-fixed effects by exploiting panel data. However, the reliability of those estimates clearly depends on the longitudinal (within-country) variation of these factors that, in the case of a persistent and structural phenomenon like corruption, is admittedly low. Moreover, panel data alone do not fully address endogeneity concerns, as a variation in corruption and in the outcome variable might reflect common (country-specific) shocks. To tackle these difficulties, we exploit a different identification strategy that hinges on the differential impact of corruption among individuals living in the same area, while controlling through area fixed effects for any other potential omitted variable correlated with corruption.

Second, the measurement of corruption itself may be problematic either from a cross-sectional or longitudinal point of view. Indeed, one may question the capacity of international surveys to capture the intensity of corruption equally well in all countries, due to differences in culture and social norms or to other perception biases. Similarly, official data on reported crimes might not be comparable across countries due to differences in laws or in the availability of harmonized crime statistics. The extent of these measurement issues can also vary over time. However, our analysis is based on various measures of corruption within a single country, thus exploiting (sub-national) territorial variability while using homogeneous and comparable indicators.

Third, previous studies on the relationship between corruption and workforce sorting have used experimental evidence and focused on whether workers' personal propensity to dis-

honesty makes them more likely to self-select into the public sector. On the contrary, we rely on hard data and drive the attention on the impact of environmental levels of corruption on sorting based on human capital, measured by individual educational attainments and grades obtained at school. Personal attitudes towards unlawful behaviors are undoubtedly of the foremost importance in determining the quality and the impartiality of public services. However, poor human capital endowments might also hinder the effectiveness of economic decisions by public agencies on a number of relevant dimensions, such as the level and the composition of public expenditure, the effectiveness of public investments and the quality of public services provided to households and firms.

Fourth, we focus on corruption's impact on both selection and allocation processes, while previous studies have mainly directed their attention towards workforce sorting only. However, human resource misallocation is also relevant: on the one hand, the same group of individuals can produce substantially different results if they are badly matched to jobs requiring different educational qualifications; on the other hand, bad allocation processes and misaligned career rewards might discourage the most skilled individuals from applying for a public job in the first place.

Finally, though we cannot observe individual performance, we are able to examine the impact of corruption on proxies of effort exerted by public employees. We find that corruption is associated to lower labor supply (in terms of hours worked) and higher absence rate, consistent with worse personnel selection and allocation processes. As a whole our findings point to a lower quality and a reduced efficiency of the public sector.

The rest of the paper is organized as follows. Section 3.2 describes the data sources and the main variables of our analysis, including the construction of corruption indicators. Section 3.3 presents our empirical strategy and clarifies which effects of corruption we are able to identify. Section 3.4 presents our main findings and some robustness checks. Section 3.5 concludes.

3.2 Data and variables

3.2.1 Individual information on occupation and schooling

Individual data on employment characteristics and observed measures of human capital are drawn from two sources. The main one is the Italian Labour Force Survey (LFS). The survey is carried out by the National Institute of Statistics (Istat) on a weekly basis

and its main aim is to provide accurate and official statistics concerning the employed and unemployed population in Italy. We pool the LFS waves from 2004 to 2010 and we restrict the analysis to non-manual employees (i.e. ISCO major groups 1 to 5).[1] LFS does not provide a clean distinction between the public and the private sector and, therefore, we identify as public employees all those employed in the following three NACE 2-digits groups: public administration, education and health. We also know the professional qualification of each employee, as measured by the ISCO occupational classification at 3 digits and their education level (in particular, the years of schooling corresponding to their highest educational attainment). Beyond the overall effect, we also provide evidence on the subgroup of managers and professionals (ISCO major groups 1 and 2), i.e. those employees who are at the top of the occupational hierarchy and who are responsible for controlling or managing an organization or staff teams. This focus is motivated by the fact that managers can single-handedly shape the activity of the public agency. We also observe labor supply and, among socio-demographic characteristics, age, gender and, most importantly for our goal, the local labor market (henceforth LLM) in which workers reside.[2] This geographic attribute is used to capture local economic and social conditions that might impact on the likelihood of joining the public sector and on the quality of the match between individual education and job skill content.

The selection process is also investigated through the use of a second data source, the Survey on Household Income and Wealth (SHIW). The survey is carried out by the Bank of Italy and contains information on the socio-economic conditions of a representative sample of the Italian population.[3] We pool the (bi-annual) SHIW waves from 2000 to 2014 and we restrict the analysis to household heads who are non-manual employees, as done with the LFS. As above, we also provide evidence on the subsample of managers and professionals. The size of the SHIW sample is much smaller than that of the LFS and details on occupation are definitely poorer. However, unlike the LFS, the SHIW allows a clean distinction between public and private sector. More importantly, the SHIW can be used to complement the LFS analysis with a further dimension of individual human capital: the final grade relative to the individual's highest educational attainment. Among

[1]The elaboration on Istat data for this work have been carried out at the Istat Elementary Data Analysis Laboratory (ADELE) in accordance with the legislation on statistical confidentiality and personal data protection. Reported results and their interpretation are to be considered sole responsibility of the authors and do not by any means represent official statistics or involve Istat in any other way. All analyses reported hereunder do not make use of sample weights.

[2]LLMs are geographic units composed of contiguous municipalities and delimited on the basis of daily commuting patterns; therefore, a LLM represents the area in which most individuals both reside and work.

[3]See Brandolini and Cannari (1994) for more details on the survey.

Table 3.1: Descriptive statistics

	Full sample		*Public sector employees*	
	Mean	St. dev.	Mean	St. dev.
		LFS data		
Female	0.561	0.496	0.656	0.475
Young (<35)	0.293	0.455	0.149	0.356
Years of schooling	12.53	3.337	13.54	3.467
Under-education	0.127	0.333	0.124	0.330
Over-education	0.090	0.286	0.103	0.304
Hours worked per week	34.859	10.348	31.883	10.013
Absence rate per week	0.015	0.121	0.019	0.136
# observations	*753,048*		*301,120*	
		SHIW data		
Female	0.365	0.482	0.382	0.486
Young (<35)	0.110	0.313	0.073	0.260
Years of schooling	13.09	3.256	13.43	3.481
Grades obtained at school	0.824	0.136	0.837	0.138
# observations	*11,511*		*5,905*	

Years of schooling are those corresponding to the highest educational attainment. Grades are the final grades (normalized with respect to the maximum obtainable grade) obtained at the highest education attainment (they are available only for individuals with at least a diploma). Under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution within his/her profession (based on the 3-digit classification). The statistics on hours worked in the last week do not include individuals that worked 0 hours. The absence indicator is 1 for those who have been absent from work for reasons including: sick leave, working hours flexibility, study leave, absence for family reasons and for undeclared causes.

Sources: authors' elaborations on data drawn from LFS and SHIW.

socio-demographic characteristics, we include – as before – age, gender and the LLM where the individual resides.

Descriptive statistics for the LFS and SHIW samples are reported in Table 3.1. Consistently with evidence from previous studies (see for instance Giorgiantonio et al., 2016, and Rizzica, 2016), in both samples public sector employees are, on average, older, include a larger share of women and possess relatively richer endowments of human capital, both in terms of education attainments and grades obtained at school. Moreover, the extent of average mismatches (both under-education and lower-education) is similar in the private

and the public sector. Finally, public employees work on average less than private employees and are slightly more likely to be absent for non-exogenous reasons (see subsection 3.4.5 for details on how the absence rate is defined) in the week prior to the survey.

3.2.2 Measures of corruption

Measuring corruption is admittedly a challenging task: as all illegal activities, corruption is mostly unobservable and, therefore, difficult to quantify. There are different possible definitions of corruption and different approaches to its measurement: so far, four basic approaches have been used to measure corruption at some aggregate level.

The first approach is based on *subjective* and *direct* assessments or perceptions of the extent of corruption, drawn from *ad hoc* surveys among citizens or "expert" respondents. Indicators of this type include, for example, Transparency International's Corruption Perceptions Index and Global Corruption Barometer, as well as the European Quality of Government Index (see Charron et al., 2014). The second approach relies on *subjective* but *indirect* indicators of corruption. For example, distrust towards (local or national) governments might at least partially reflect their perception as corrupt entities. The third approach relies on *objective* though *indirect* measures of corruption. For example, "missing expenditure" (Olken, 2009) – i.e. the difference between public expenditure in certain infrastructures and the corresponding realized outcome – can be thought as an observable consequence of corruption. The fourth approach relies on *objective* and *direct* measures of corruption such as direct observations, reported crimes or similar evidence arising from government audits.[4]

Each of these approaches has advantages and drawbacks. Subjective and perception-based indicators (either direct or indirect) have been widely used, as they are available for a large set of countries and allow to exploit cross-country variation in corruption to examine the latter's relationship with other economic outcomes.[5] However, the effectiveness of these indicators has been questioned. First, there are significant differences in cultural traits, social norms and laws across countries, so that citizens of one polity may

[4]Olken and Barron (2009) designed a study in which surveyors accompanied Indonesian truck drivers on their trips in order to collect direct observations on illegal payments to police, soldiers, and weigh station attendants. Ferraz and Finan (2011) and Brollo et al. (2013) used data on a program of random audits on local governments, with detailed reports on corruption charges. For Italy, Del Monte and Papagni (2001, 2007) and Barone and Mocetti (2014) used official statistics on reported crimes against the public administration.

[5]See, among the others, Mauro (1995), Knack and Keefer (1995), La Porta et al. (1999), Fisman and Gatti (2012) and Fisman and Miguel (2007).

find certain practices more acceptable than citizens of another, thus leading to different reported perceptions of the extent of corruption. Second, the reliability of survey information has also been questioned, as respondents might not report direct experiences but be influenced by what is publicized in the media (Rizzica and Tonello, 2015). The third approach is also intriguing, but missing expenditure – like any other variable measured as a "residual" – is not necessarily attributable to corruption. For example, the effectiveness of public spending in infrastructures might also reflect the efficiency of the local construction industry, unobserved characteristics of the territory or other random elements, thus confounding the interpretation of the computed indicator. Finally, the fourth approach, beyond poor cross-country comparability due to differences in laws and in the organization of the judicial system, might suffer from reporting bias. If crime episodes are collected by police forces or courts, variations in their number might reflect not only the intensity of the criminal activity, but also the efficiency of such institutions and/or their interest in prosecuting that particular type of offense.

In this paper we adopt two measures of corruption. The first is based on crimes reported by police forces to the judicial authority, extracted from the SDI database.[6] Data at our disposal are collected at the municipality level and cover the period from 2004 to 2011. In particular, we restrict the analysis to crimes intimately linked to corruptive practices: corruption proper, graft and malfeasance.[7] These unlawful behaviors all result into additional payoffs accruing to the public employee at the detriment of one or more private agents: the role of such agents may range from being an active part in the enactment of the criminal deed (as in bribery) to being the victims of the public servant's prevarication (as in graft). These raw figures on crimes are normalized with respect to total employment at the local level (a proxy for the level of economic transactions). This measure is computed at the local labor market (LLM) level and is averaged over the period of observation.[8] To address potential reporting bias, we partial out the effect of the local

[6]SDI (*Sistema Di Indagine*) is managed by the Ministry of Interior and collects data on crimes reported by the three main Italian police forces (*Arma dei Carabinieri*, *Polizia di Stato* and *Guardia di Finanza*).

[7] Crimes perpetrated by public officials are regulated by the Italian criminal law (*Codice Penale*, articles 314-323, 479-481 and 493): acknowledging oversimplification, *corruption* ('*corruzione*' in Italian) proper takes place when the public official accepts a bribe from a private counterparty in exchange for the enactment of or the abstention from certain behaviors; *graft* ('*concussione*') refers to the situation in which the payment is imposed by the civil servant to the private party; here *malfeasance* ('*abuso d'ufficio*') generically defines behaviors enacted by the public employee aiming at earning unlawful benefits: resource embezzlement ('*peculato*') and document forgery ('*falsità ideologica*'), when perpetrated by public officials or by other providers of public services, may be seen as special cases of malfeasance.

[8]We do not exploit within-LLM variation since corruption is a persistent phenomenon and does not show sufficient longitudinal variation.

judicial efficiency on crime rates. Namely, we run a regression where we control for the judicial efficiency (as measured by the lengths of penal proceedings in local courts) and we take the residuals. The latter yield a measure of corruption incidence net of local judicial efficiency (C^1 henceforth).[9]

Table 3.2: Corruption: principal component analysis

	1st component	2nd component	3rd component	4th component
Eigenvalue	2.573	0.883	0.330	0.214
Proportion	0.643	0.221	0.082	0.054
Cumulative	0. 643	0.864	0.946	1.000
	C^1	Trust	GP	CPI
Coefficient 1st component	0.365	0.482	0.382	0.486

The second measure is a synthetic indicator that combines information drawn from different approaches. Namely, we collect four different variables, each echoing one of the four measurement methods mentioned above (though, for reasons of data availability, they are measured at different geographical levels). The first variable is a subjective assessment of the level of corruption (CPI). Data are drawn from a large European survey (EQI) aimed at measuring the quality of governance within the European Union and they are available at the regional level.[10] The second variable echoes the subjective and indirect approach to the measurement of corruption. We exploit a survey managed by Istat (the so-called "Multiscopo") asking a large set of questions to citizens to various aspects of life, including trust towards local government and other institutions ($TRUST$). Based on a rich literature on the detrimental effect of perceived corruption on the trust expressed by people in local institutions (e.g. Uslaner, 2004; Clausen et al., 2011), we take distrust towards local government as an indicator of corruption. These figures are available at the regional level with a further distinction between small municipalities, intermediate municipalities and larger metropolitan areas. The third variable belongs to the group of objective and indirect measures of corruption. Golden and Picci (2005) compute a measure of corruption for Italy based on the difference between the value of the public infrastructure and cumulated public expenditure in public works (GP). These figures are available at the regional level. Our last variable is reported crime adjusted for judicial

[9]According to our findings, a variation of one standard deviation of the length of penal proceedings is associated to a 0.14 standard deviations increase in the reported crime rate, thus suggesting that the latter largely reflects the intensity of the criminal activity at the local level and is only marginally affected by judicial efficiency (as captured by our proxy).

[10]More information on the data can be found in Charron et al. (2014).

efficiency, i.e. the aforementioned C^1. We then rely on a principal component analysis to extract information from these four variables. The first principal component explains about 64 percent of the total variance of the underlying variables and it is positively associated, as expected, with each one of the input variables (Table 3.2). We call this synthetic indicator C^2.

C^1 and C^2 both have some strong and weak points. On the one hand, C^2 might better capture a multidimensional and unobserved phenomenon such as corruption; moreover, the large fraction of variance explained by the first component suggests that the four indicators largely overlap, which is supportive of the measure's rich informational content. On the other hand, C^1 is easier to interpret in economic terms and less subject to arbitrary choices. Moreover, C^1 is available at a finer partition of the territory, while C^2 partly reflects indicators that mostly vary at the regional level. For these reasons, C^1 is our preferred measure of corruption, though we provide evidence using both indicators throughout the paper.[11]

Table 3.3: Corruption: descriptive statistics

	Mean	S.D.	10th	25th	50th	75th	90th
Crime rate: C^1	0.000	1.000	-0.570	-0.442	-0.254	0.141	0.765
Principal component: C^2	0.000	1.000	-1.299	-0.739	0.133	0.856	1.225

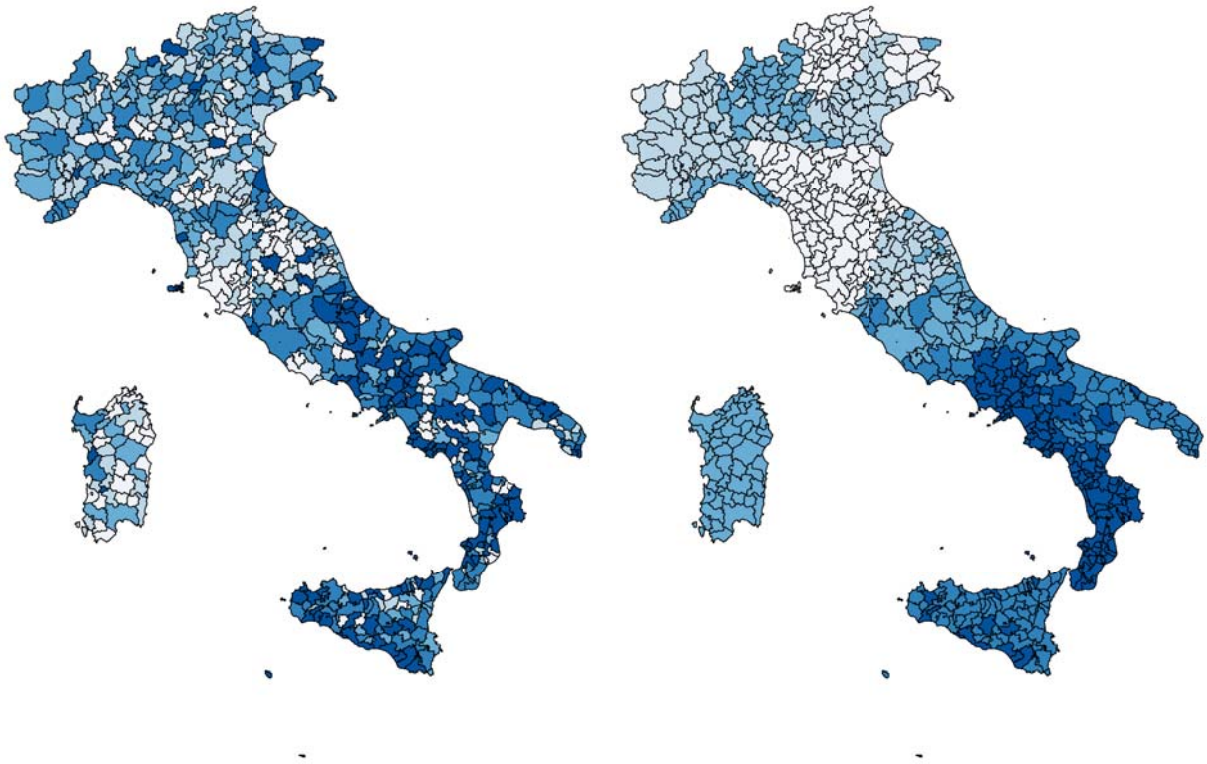
Corruption indicators are standardized at the LLM level.

Summary statistics of the two indicators are reported in Table 3.3. In order to guarantee comparability between different indicators of corruption, both C^1 and C^2 are standardized. The two variables display considerable variability across LLMs. A graphical representation of the territorial differences in corruption intensity is reported in Figure 3.1: both indicators show that corruption is more widespread in Southern Italy, with the North-South divide being more visually evident when C^2 is used; however, in both cases there is also significant variability within each macro-area.

3.2.3 Descriptive evidence

Corruption and human capital endowments are positively correlated at the LLM level, as shown in Figure 3.2. This apparently surprising fact is mainly due to other covariates being correlated to both variables. For example, corruption is positively related to the size

[11]Notice that our main results are qualitatively confirmed even if we use raw figures for reported crimes and/or each component of the principal component analysis separately. Results are available from the authors upon request.



(a) Crime rate: C^1. (b) Principal component: C^2.

Figure 3.1: A map of corruption at the LLM level.

of the public sector (Figure 3.3), either because large public agencies offer better chances to corruptors or because corruption may hinder the development of more market-oriented activities. Sector composition, however, also affects the incentives to invest in human capital: as it has been widely documented, the public sector tends to attract the most educated workers (Cowley and Smith, 2014; Rizzica, 2016). Moreover, corruption is more widespread in less economically developed LLMs, as measured by the value added per capita, and by poorer labor market opportunities. The latter, however, might also affect human capital investments reducing *ceteris paribus* the opportunity cost of studying. Indeed, when we control for the above-mentioned variables, the correlation between corruption and education disappears. This also suggests that, when attempting to identify a clean effect of corruption on other socio-economic outcomes, we face the challenging task of having to avoid spurious correlation driven by unobserved omitted variables.

Having established that the share of employees in the public sector is larger in LLMs where the intensity of corruption is higher, we found that the share of managers among public employees in those areas is also larger (31% versus 27%, see Figure 3.3). In terms of human

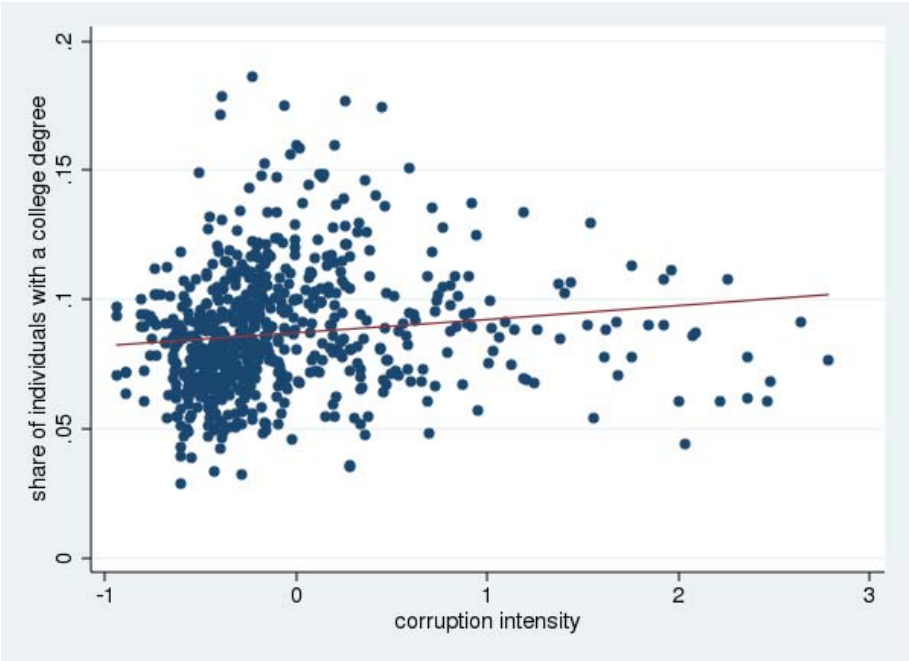


Figure 3.2: Corruption and education across LLMs. Corruption intensity is measured by C^1; the share of population with a college degree is drawn from the 2001 Census.

capital, as measured by the years of schooling, the (positive) gap between the public sector and the private sectors is narrower in more corrupt LLMs, and this is especially evident for managers, for whom average schooling advantage falls from 2.1 to 1.6 years (Figure 3.4). Moreover, the relationship between under-education in the private and public sector changes when one moves from less to more corrupt areas (Figure 3.5). When corruption is low, public-sector employees are *less* likely to be under-educated than private-sector employees by 2.4 percentage points (1.8 points for managers). On the contrary, where corruption is high, public-sector employees are *more* likely to be under-educated than private-sector employees by 1.1 percentage points (2.7 points for managers).

3.3 Empirical strategy

3.3.1 Selection of workers into the public sector

The first phenomenon we wish to study is the potential distortionary effect of corruption on the relevance of educational attainments, as well as other measures of individual ability,

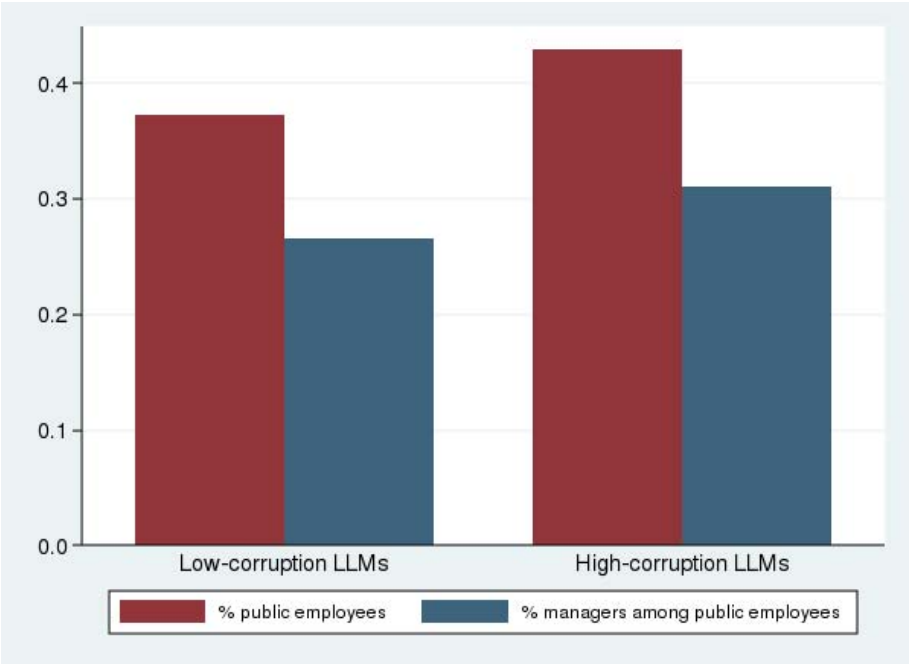


Figure 3.3: Share of public employees over working population and share of managers among public employees by corruption intensity.

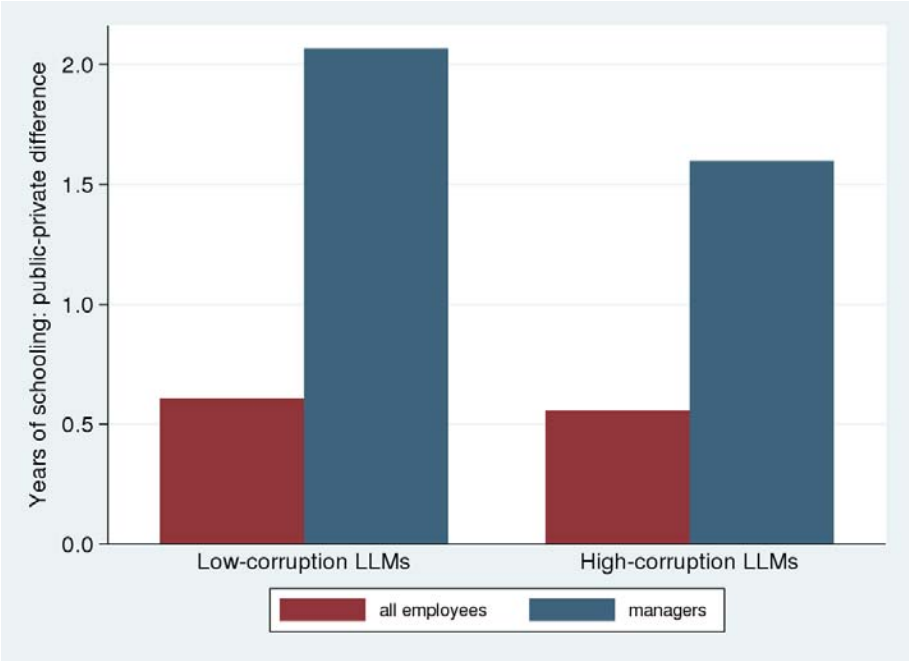


Figure 3.4: Public-private differences in schooling.

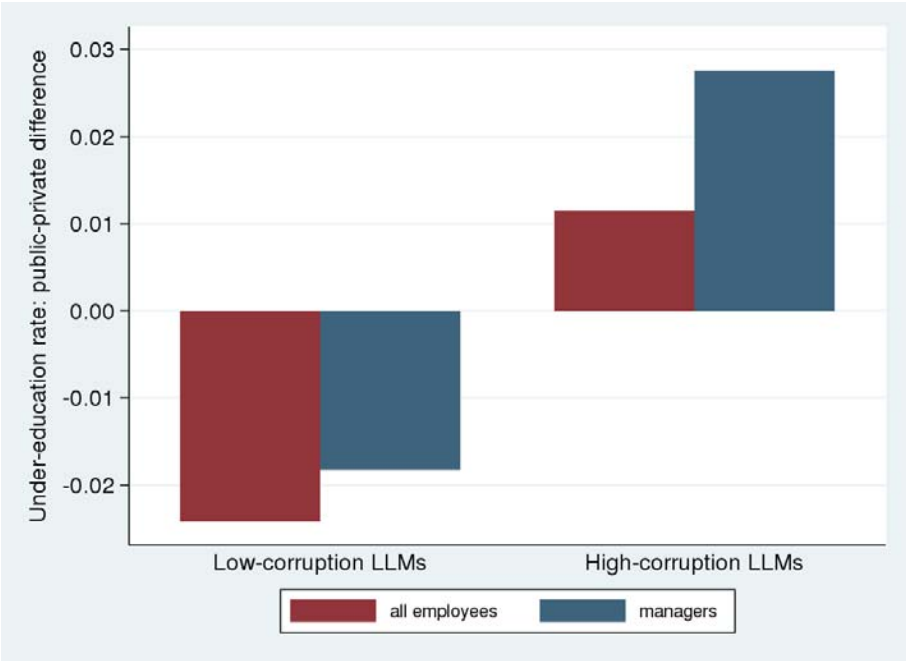


Figure 3.5: Public-private differences in under-education.

as predictors of the likelihood of an individual's belonging to the public sector. To this end, we estimate the following linear probability model:

$$Y_i = \alpha + \beta S_i + \delta \left(S_i \cdot C_{LLM(i)} \right) + \gamma' X_i + \rho_{LLM(i),p(i)} + \varepsilon_i \tag{3.1}$$

where Y_i is a binary indicator of the occupational status of individual i, taking on the value of 1 if i is a public employee and the value of 0 if i is employed in the private sector; S_i is a measure of i's skills endowment, e.g. the completed years of schooling; $C_{LLM(i)}$ is one of the two measures of the incidence of corruption in the LLM in which individual i resides; X_i is a vector of individual controls such as gender and age: these are included as the likelihood of joining the public sector may be affected by gender- or cohort-specific factors; finally, the term $\rho_{LLM(i),p(i)}$ is a group indicator, obtained by combining i's LLM and professional area, capturing unobserved factors at that level.[12] Thus, our coefficient of interest δ captures how the impact of schooling on the likelihood of joining a certain professional class $p(i)$ in the public sector (rather than the same professional class in the private sector) varies across LLMs characterized by different corruption intensity. In most specifications, we compare the public sector with the manufacturing sector, rather than

[12] In RFL data we can distinguish, within the sample of non-manual employees, among three professional areas: managers and professional, technicians, and clerical workers. In SHIW data we only distinguish between managers and clerical workers.

with the entire private sector. The former, being hardly dependent on public spending and more exposed to international competition, tends to be more-market oriented and, therefore, is assumed to be substantially unaffected by corruption.[13]

We might expect $\delta < 0$, indicating that corruption reduces the likelihood to join the public sector.

3.3.2 Allocation of workers within the public sector

Besides affecting, through self-selection and screening, the composition of the available public workforce, corruption may have an impact on how efficiently these human resources are assigned to different jobs and tasks. In particular, we imagine that to each job there corresponds a level of skills or human capital, i.e. the level of an individual which is "just right" for that job. We subsequently test whether corruption shifts the allocation of human resources away from the right matching and, if that happens, whether this prevalently takes the form of under- or over-education, i.e. of employees having a much lower or higher, respectively, skill level than that required on average by the jobs they are assigned to. In order to quantify this effect, we estimate the following linear probability model:

$$M_i = \alpha + \beta \left(Y_i \cdot C_{LLM(i)}\right) + \gamma' X_i + \rho_{LLM(i)} + \varphi_{s(i)} + \varepsilon_i \tag{3.2}$$

where the dependent variable M_i is a binary indicator for the presence of some form of skills mismatch (under-education or over-education) for individual i. Specifically, an individual is considered to be under-educated if her schooling level falls below the 25th percentile of the distribution of schooling within her profession (defined in terms of the ISCO classification at 3 digits) and, conversely, over-educated if her schooling level exceeds the 75th percentile of that distribution. As before, Y_i denotes whether i is employed in the public rather than in the manufacturing sector. LLM-fixed effects ($\rho_{LLM(i)}$) and sector-fixed effects ($\varphi_{s(i)}$) capture local or industry-specific variables that might be correlated with mismatch.

Our coefficient of interest is β, which captures how the impact of working in the public sector on the likelihood to be mismatched varies across LLMs characterized by different corruption intensity. We might expect $\beta > 0$, as corruption is supposed to increase

[13] Indeed, corruption typically involves converging interests or – at least – some kind of interaction between the public officials and the private firms whose activity is affected by public decisions (e.g. electricity, water, waste disposal, construction sectors, social activities, etc.). However, in the robustness section we relax this assumption and we allow the employee to choose also between the public sector and the overall private sector.

mismatch in the public sector relative to the manufacturing sector.

3.3.3 Identification assumptions

When we examine the impact of corruption on economic outcomes exploiting cross-sectional variation, we should take account of two potential identification threats.

First, unobserved heterogeneity at the local level (e.g. social norms, level of economic development, etc.) might be related to corruption as well as to the accumulation of human capital. These omitted variables are likely to bias the OLS estimates. However, we include LLM-fixed effects aimed at capturing any potential variables at the local level. Indeed, our identification strategy exploits the differential effect of corruption between individuals living in the same LLM (e.g. between those employed in the public or in the private sector). This strategy mimics a difference-in-differences approach where the treatment is represented by the intensity of corruption at the local level and the individual characteristics of the employees determine the exposure to the treatment.

Second, we might suspect the presence of reverse causality, as one may argue that skill-biased recruitment and human resources management processes in the public sector could affect the intensity of corruption. To address this problem, we exploit variation in corruption intensity at the local level that is attributable to factors associated to corruption *but* predating the hiring of the current public employees. In particular, we use two types of pre-determined indicators as instruments. First, we exploit variation in corruption that is attributable to economic rents at the local level. More specifically, we use data from the Italian 1971 Census to compute the dependence of the private sector from public demand.[14] The idea is that where the latter is higher, the economic rents associated to the discretionary power of the public officials as well as the incentive of the entrepreneurs to influence public spending are also higher. This, in turn, may have increased the likelihood of corruptive practices being established. Second, we exploit variation in corruption that is attributable to historical dominations by the following nations or families: Anjou, Austria, Bourbons, Normans, Papal States, Savoy, Spain, Swabians and the Republic of Venice (with the independent states being the residual category).[15] This analysis is related to a large literature that investigates how history (and historical institutions) may still

[14]Dependence on the demand of the public sector at the local level is computed in two steps. First, using the input-output matrix, we compute the dependence on the public demand for each sector of economic activity. Second, we translate these figures at the local level using the past sector composition of the local economy (i.e. the distribution of employees across sectors at the local level as recorded by the 1971 Census).

[15]Figures are drawn from Di Liberto and Sideri (2015).

influence existing institutions and current social behaviors (e.g. Acemoglu and Robinson, 2012).

Table 3.4: Selection in the public sector: the impact of schooling (LFS)

Dependent variable:	Employed in the public sector			
Professional area:	All	Managers/ professionals	All	Managers/ professionals
Years of schooling	0.016***	0.025***	0.015***	0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
Years of schooling × C^1	-0.008***	-0.013***		
	(0.001)	(0.002)		
Years of schooling × C^2			-0.007***	-0.009***
			(0.001)	(0.001)
LLM × professional area FEs	YES	YES	YES	YES
R-squared	0.269	0.218	0.270	0.220
# observations	*397,060*	*99,663*	*397,060*	*99,663*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variable is equal to 1 for public sector employees and to 0 for manufacturing sector employees. Years of schooling are those corresponding to the highest educational attainment. Corruption is measured at the LLM level and we consider two measures: C^1 – i.e. reported crimes net of judicial efficiency – and C^2 – i.e. the principal component of CPI, $TRUST$, GP and C^1. Other controls include fixed effects for gender and age cohort.

3.4 Results

3.4.1 The impact of corruption on personnel selection

Table 3.4 reports the results of the estimation of model (3.1) for our two main measures of corruption, C^1 and C^2. The sample, drawn from the LFS, include all the employees of the public and manufacturing sectors engaged in non-manual activities. Individual human capital is measured by years of schooling.

Higher educational attainments are, as expected, positively associated with the likelihood of having joined the public sector. One additional year of schooling increases the probability of being a public employee by around 1.6 percentage points; the impact is higher among managers and professionals (2.5 percentage points). More interestingly, corruption reduces the role of education as a predictor of being a public employee. According to our

results, moving from a LLM at the 10th percentile of C^1 to one at the 90th percentile (i.e. from a low-corruption to a high-corruption LLM) the impact of one additional years of schooling decreases from 2.1 to 1.0 percentage points; the detrimental effect of corruption is larger for managerial and professional occupations, where the same exercise would lead to a decrease of the impact of education from 3.2 to 1.5 percentage points. The last two columns of Table 3.4 replicates the analysis using C^2 as an approximation of corruption intensity at the local level: results are qualitatively similar.

In Tables 3.5 and 3.6 we rely on the SHIW data rather than on the LFS. Results should be interpreted with some caution given the relatively small number of observations and the large number of fixed-effects that we include in the specification in order to control for the relevant unobserved heterogeneity. However, SHIW data allow us to use a second measure of ability, i.e. an index representing the grade obtained by individuals at their highest achieved educational level, which is available only for those with at least secondary education.

Table 3.5: Selection in the public sector: the impact of schooling (SHIW)

Dependent variable:	Employed in the public sector			
Professional area:	All	Managers/ professionals	All	Managers/ professionals
Years of schooling	0.005***	0.006***	0.005***	0.006***
	(0.002)	(0.002)	(0.001)	(0.001)
Years of schooling × C^1	-0.003	-0.004*		
	(0.003)	(0.002)		
Years of schooling × C^2			-0.003**	-0.003**
			(0.002)	(0.001)
LLM × professional area FEs	YES	YES	YES	YES
R-squared	0.203	0.185	0.204	0.186
# observations	*4,939*	*2,419*	*4,939*	*2,419*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the SHIW. The dependent variable is equal to 1 for public sector employees and to 0 for industrial sector employees. Years of schooling are those corresponding to the highest educational attainment. Corruption is measured at the LLM level and we consider two measures: C^1 – i.e. reported crimes net of judicial efficiency – and C^2 – i.e. the principal component of *CPI*, *TRUST*, *GP* and C^1. Other controls include fixed effects for gender and age cohort.

In Table 3.5 we consider years of schooling as the only ability measure and the full sample of (non-manual) employees. These results generally confirms previous ones, though the

impact of education on the probability of joining the public sector is, on average, slightly weaker. Estimates for our parameter of interest, albeit only weakly significant, testify the presence of a detrimental effect of corruption, again concentrated among managers.

Table 3.6: Selection in the public sector: the impact of grades (SHIW)

Dependent variable:	Employed in the public sector			
Professional area:	All	Managers/ professionals	All	Managers/ professionals
Grades at school	-0.010**	0.018	-0.010**	0.015
	(0.046)	(0.040)	(0.045)	(0.039)
Grades at school × C^1	0.037	-0.110*		
	(0.052)	(0.059)		
Grades at school × C^2			-0.006	-0.077***
			(0.028)	(0.024)
LLM × professional area FEs	YES	YES	YES	YES
R-squared	0.207	0.185	0.206	0.187
# observations	*4,926*	*2,414*	*4,926*	*2,414*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the SHIW. The dependent variable is equal to 1 for public sector employees and to 0 for industrial sector employees. Grades are the final grades (normalized with respect to the maximum obtainable grade) obtained at the highest education attainment (they are available only for individuals with at least a diploma). Corruption is measured at the LLM level and we consider two measures: C^1 – i.e. reported crimes net of judicial efficiency – and C^2 – i.e. the principal component of CPI, $TRUST$, GP and C^1. Other controls include years of schooling and fixed effects for gender and age cohort.

In Table 3.6 we restrict the analysis to individuals with at least a secondary education diploma and we focus the attention on ability as measured by school grades.[16] According to our findings, having obtained one additional grade-point, in a scale ranging from 0 (the lowest grade) to 10 (the highest grade), has a negative effect on the overall probability of joining the public sector (with respect to the manufacturing sector). However, the impact is positive (though not statistically significant) when we focus on managers and other professionals. Again the impact of grades is differentiated across LLMs characterized by a different intensity of corruption, which negatively affects the propensity of more talented students to join the public sector. According to our estimates, one additional grade-point increases the likelihood of joining the public sector in a managerial position

[16]Grades are highly correlated with educational attainments, as those who obtain higher grades at secondary level are also those who are more likely to get tertiary education. Therefore, to avoid collinearity, we do not consider the two ability measures jointly interacted with corruption.

by 8 percentage points in low-corruption LLMs and decreases it by 6 percentage points in high-corruption LLMs. Results are again qualitatively similar if we use C^2 instead of C^1 to measure corruption.

3.4.2 The impact of corruption on personnel allocation

In this section we inspect the impact of corruption on the effectiveness of the allocation process of human resources. The latter is examined comparing individual abilities and the skill content of jobs workers are assigned to. A mismatch may happen both in the direction of under-education (an individual is assigned to a task which is on average undertaken by more educated workers) or over-education (an individual is assigned to a task which is on average undertaken by less educated workers).

Table 3.7: Under- and over-education in the public sector

Dependent variable:	Under-education		Over-education	
Professional area:	All	Managers/ professionals	All	Managers/ professionals
Public sector × C^1	0.041***	0.041***	0.002	0.001
	(0.007)	(0.013)	(0.005)	(0.008)
LLM FEs	YES	YES	YES	YES
Sector of activity FEs	YES	YES	YES	YES
R-squared	0.045	0.059	0.043	0.038
Public sector × C^2	0.018***	0.015**	0.001	-0.004
	(0.003)	(0.006)	(0.003)	(0.004)
LLM FEs	YES	YES	YES	YES
Sector of activity FEs	YES	YES	YES	YES
R-squared	0.045	0.059	0.043	0.038
# observations	*397,064*	*99,663*	*397,064*	*99,663*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variable under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (ISCO occupational at 3 digits). Years of schooling are those corresponding to the highest education attainment. Corruption is measured at the LLM level and we consider two measures: C^1 – i.e. reported crimes net of judicial efficiency – and C^2 – i.e. the principal component of *CPI*, *TRUST*, *GP* and C^1. Other controls include fixed effects for gender and age cohort.

As we have described before, under- and over-education are by and large as frequent in the public sector as they are elsewhere. The aim of our empirical strategy is, again, to examine differential patterns between low- and high-corruption areas. Table 3.7 shows

the results of the estimation of model (3.2). The coefficient associated to the interaction term between the public sector dummy and the measure of corruption is positive, suggesting that corruption increases the correlation between being in the public sector and the likelihood to be under-educated. These results hold for all employees and for the subset of those who stay at the top of the occupational hierarchy. On the other hand, we do not find any detectable effect in terms of over education; this might also be due to an inflation of professions with a higher (formally required) skill content in more corrupt LLMs, thus making over-education less likely by definition.

Table 3.8: Under- and over-education (education and skill content corrections)

Dependent variable:	Under-education		Over-education	
Professional area:	All	Managers/ professionals	All	Managers/ professionals
Public sector × C^1	0.026***	0.025*	0.011**	0.007
	(0.006)	(0.013)	(0.005)	(0.008)
Average skill content	0.083***	0.068***	-0.054***	-0.027***
	(0.003)	(0.007)	(0.003)	(0.004)
Average educational level	-0.076***	-0.069***	0.059***	0.026***
	(0.002)	(0.005)	(0.002)	(0.004)
LLM FEs	YES	YES	YES	YES
Sector of activity FEs	YES	YES	YES	YES
R-squared	0.054	0.064	0.050	0.040
# observations	*397,064*	*99,663*	*397,064*	*99,663*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variable under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (ISCO occupational at 3 digits). Years of schooling are those corresponding to the highest education attainment. Average skill content is the LLM-by-sector is measured as the product of the nation-wide average of schooling in each profession and the share of each professions in the sector-LLM cell; average educational level is the average schooling of the employed in the sector-LLM cell. Corruption is measured at the LLM level with C^1, i.e. reported crimes net of judicial efficiency. Other controls include fixed effects for gender and age cohort.

Skills mismatch might be, at least partially, a mechanical consequence of the negative selection patterns observed in the previous subsection. If corruption makes public employment relatively less attractive for the most educated, public agencies in corrupt areas will hire relatively less educated personnel. Assuming that the tasks assigned to each agency do not vary with the level of corruption, under-education will arise as the obvious outcome of having to fill the same job positions with less educated personnel. But under-education could also result from biased management practices, that may be more likely to

occur where corruption is more intense. As an attempt to disentangle these two factors, in Table 3.8 we estimate a model identical to (3.2) but for two additional controls: the average skill content of professions present in each sector-LLM cell (measured as product of the nation-wide average of schooling in each profession and the share of professions in the cell) as well as the average education level in the same sector-LLM cell (measured with the average schooling of the employed in the cell).[17] These should respectively account for different educational endowments (and thus for the effects of selection) as well as for possible inflation in the number of high-level positions managed by public agencies. We indeed find, as expected, that these additional controls partly explain the levels of under- and over-education; for example, under-education is more likely where the average schooling of employees is lower and where the average schooling required by the available job positions is higher. Nevertheless, we still find evidence of corruption-related variations in the likelihood of under-education.

3.4.3 Robustness

This section contains some robustness checks, motivated by various considerations.

First, we examine whether our results hold when we modify the control group or use different sample selection rules. So far we have used the manufacturing sector as control group as it is hardly dependent on public spending and, therefore, arguably unaffected by corruption in the public sector. In the first two columns of Table 3.9, however, we replicate our baseline results on selection, and under- and over-education using C^1 as our measure of corruption and extending the control group from the manufacturing to the whole private sector.[18] Our main results are fully confirmed though the estimated effects are slightly smaller, thus implicitly suggesting that misallocation due to corruption is somewhat extended also to the private industries that interact more with the public sector.[19] Moreover, having shown that corruption is more widespread in the South of

[17] This is the LLM-by-sector average of the average education level within professions at the ISCO 3-digit level of disaggregation.

[18] All results of this subsection are qualitatively similar if we use C^2 instead of C^1. They are not reported for the sake of brevity.

[19] One can also argue that a sharp comparison between public and private sector may fail to take into account that the latter is also potentially affected by corruption and possibly differently so depending on the industry. Indeed, corruption typically involves converging interests or – at least – some kind of interaction between the public officials and the private firms whose activity is affected by public decisions. In unreported evidence we replicate the analysis using a continuous indicator of dependence from the public sector in lieu of the discrete indicator. More specifically, we map economic activities into the unit interval, capturing the dependence and/or the proximity between each economic sector of activity and the

Table 3.9: Robustness: different control groups

Dependent variable:	Employed in the public sector			
Control group:	All private sector		Only Centre-North	
Professional area:	All	Managers/ professionals	All	Managers/ professionals
Years of schooling	0.022***	0.045***	0.018***	0.030***
	(0.001)	(0.001)	(0.001)	(0.002)
Years of schooling × C^1	-0.002**	-0.009***	-0.012***	-0.017***
	(0.001)	(0.002)	(0.004)	(0.006)
LLM × professional area FEs	YES	YES	YES	YES
R-squared	0.255	0.280	0.253	0.229
# observations	*753,043*	*135,127*	*263,176*	*58,995*
Dependent variable:	Under-education			
Public sector × C^1	0.009**	0.021***	0.057***	0.078***
	(0.004)	(0.007)	(0.014)	(0.027)
LLM FEs	YES	YES	YES	YES
Sector of activity FEs	YES	YES	YES	YES
R-squared	0.058	0.060	0.050	0.058
Dependent variable:	Over-education			
Public sector × C^1	0.000	-0.000	0.009	0.031
	(0.003)	(0.003)	(0.011)	(0.019)
LLM FEs	YES	YES	YES	YES
Sector of activity FEs	YES	YES	YES	YES
R-squared	0.048	0.039	0.047	0.040
# observations	*753,048*	*135,127*	*263,180*	*58,995*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the LFS. As control groups, in the first two columns all employees in the private sector are included while in the last two columns only employees located in the Centre-North of Italy are included. The dependent variables are the following: employed in the public sector is equal to 1 for public sector employees and to 0 for private sector employees; under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (ISCO occupational at 3 digits). Years of schooling are those corresponding to the highest education attainment. Corruption is measured at the LLM level with C^1, i.e. reported crimes net of judicial efficiency. Other controls include fixed effects for gender and age cohort.

public sector, using the input-output matrix. The smallest values of this continuous measure correspond

Italy, we examine to what extent our results are driven by the traditional North-South divide and whether they still hold when we compare more homogeneous regions. The last two columns of Table 3.9 replicate our baseline results restricting the analysis to the LLMs located in the Centre-North, obviously at the cost of losing a significant number of observations and territorial variability. The estimates of the coefficients associated with the interaction term in the selection equation (3.1) are fairly similar to those of our baseline specification. As far as misallocation is concerned, corruption continues to be significantly associated to under-education.

Second, our results might also be driven by other omitted variables correlated with corruption and implying differential effects similar to those produced by corruption. More precisely, this concern is not related to potential omitted variables driving the sorting between public and private sectors: those are already controlled for by the introduction of fixed effects at the LLM level. The concern relates to variables having a differential schooling-biased effect similar to that observed for corruption. To address this point, we enrich the specification with other local controls aimed at capturing relevant economic dimensions that are both correlated with corruption and potentially liable to affect individuals occupational choices. In the first column of Table 3.10 we include the (logarithm of the) value added per employee interacted with schooling as a determinant of selection into the public sector. The underlying idea is that better, on average, economic prospects at the local level (and any other variable correlated with economic development) might affect the education-based sorting between public and private sector. In the second column of Table 3.10 we include population density at the LLM level: this might affect both corruption and selection patterns, since the scope of public administration can differ between urban and rural areas. Finally, in the last column of Table 3.10, we include the average unemployment rate (over the years 2004-2010) at the LLM level: unemployment and corruption may be related through a number of channels and unemployment might affect the composition of the workforce willing to join the public sector. Our main findings are basically unchanged in all cases.

Finally, we examine whether our patterns on the skill-biased impact of corruption on selection processes vary across different sections of the public sector. In Table 3.11 we replicate our baseline result distinguishing between public administration, education and health. We find that the association between schooling and the likelihood of joining the

to sectors that do not interact with the public sector (e.g. the manufacturing sector); in contrast, larger values correspond to industries whose demand partly depends on public spending and/or that operate on regulated markets (e.g. electricity, water, waste disposal, construction, etc.). The results (available upon request) are qualitatively similar.

Table 3.10: Robustness: adding further controls

Dependent variable:	Employed in the public sector		
Professional area:	All		
Years of schooling	0.016***	0.016***	0.016***
	(0.001)	(0.001)	(0.001)
Years of schooling × C^1	-0.007***	-0.008***	-0.003***
	(0.001)	(0.001)	(0.001)
Years of schooling × *VA*	0.002***		
	(0.001)		
Years of schooling × Density		0.000	
		(0.001)	
Years of schooling × Unemployment			-0.006***
			(0.001)
LLM × professional area FEs	YES	YES	YES
R-squared	0.269	0.269	0.270
# observations	*397,060*	*397,060*	*397,060*
Professional area:	Managers/ professionals		
Years of schooling	0.025***	0.025***	0.025***
	(0.001)	(0.001)	(0.001)
Years of schooling × C^1	-0.012***	-0.013***	-0.006***
	(0.002)	(0.002)	(0.001)
Years of schooling × *VA*	0.005***		
	(0.001)		
Years of schooling × Density		0.002	
		(0.003)	
Years of schooling × Unemployment			-0.009***
			(0.001)
LLM × professional area FEs	YES	YES	YES
R-squared	0.220	0.219	0.223
# observations	*99,663*	*99,663*	*99,663*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variable is equal to 1 for public sector employees and to 0 for manufacturing sector employees. Years of schooling are those corresponding to the highest educational attainment. Corruption is measured at the LLM level with C^1, i.e. reported crimes net of judicial efficiency. We also include controls that might affect the public-private sector sorting between individuals with different educational level: the (logarithm of the) value added per employee at the LLM level (*VA*), the LLM population density (Density) and the LLM unemployment rate (Unemployment), all of which are standardized. Other controls include fixed effects for gender and age cohort.

Table 3.11: Robustness: different sections of the public sector

Dependent variable:	Employed in the public sector		
Professional area:	All		
Section of the public sector:	Public administration	Education	Health
Years of schooling	0.022***	0.028***	0.012***
	(0.001)	(0.001)	(0.002)
Years of schooling × C^1	-0.009***	-0.008***	-0.001
	(0.001)	(0.002)	(0.002)
LLM × professional area FEs	YES	YES	YES
R-squared	0.359	0.509	0.310
# observations	*184,113*	*207,016*	*197,760*
Professional area:	Managers/ professionals		
Section of the public sector:	Public administration	Education	Health
Years of schooling	0.022***	0.035***	0.055***
	(0.002)	(0.002)	(0.002)
Years of schooling × C^1	-0.010***	-0.019***	-0.006**
	(0.003)	(0.002)	(0.003)
LLM × professional area FEs	YES	YES	YES
R-squared	0.355	0.318	0.390
# observations	*28,610*	*65,055*	*30,850*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variable is equal to 1 for public sector employees and to 0 for manufacturing sector employees. Years of schooling are those corresponding to the highest educational attainment. Corruption is measured at the LLM level with C^1, i.e. reported crimes net of judicial efficiency. Other controls include fixed effects for gender and age cohort.

public sector remains weaker in high-corruption areas for the three sections. Moreover, in all cases the negative impact of corruption is larger for jobs at the top of the occupational ladder.

3.4.4 IV estimates

One last concern is related to reverse causality. Corruption might itself be the result of poorly selected public employees, while we are interested in the link *from* corruption *to* personnel selection and allocation. To address this issue we rely on an instrumental variable strategy. Our instruments are characterized by the common characteristic of

being pre-dated with respect to the hiring of current public employees.

The first instrument is past local dependence from public-sector demand. The underlying idea, implicitly supported by our previous findings, is that corruption episodes are more likely to occur where the role of public spending for the private sector is more prominent. Unreported evidence documents that past economic dependence of the private sector on public demand is positively correlated with corruption intensity at the LLM level. The second set of instruments considers the impact of past dominations. Further unreported evidence shows broad consistency with other results on the cultural and institutional legacy of past foreign dominations.[20]

IV estimates are reported in Table 3.12. Past economic dependence on public spending appears to be a strong determinant of corruption and the first stage F-statistic of the excluded instrument for the whole sample is well above the traditional threshold (first two columns). On the contrary the predictive power of past dominations, in our empirically setting, is somewhat weaker and the F-statistic is, in some cases, slightly below 10 (last two columns). The second stage coefficients using either set of instruments, are qualitatively similar to those of our baseline specifications, which reassures us on the identification of a link from corruption to personnel selection and allocation in the public sector.

3.4.5 The impact of corruption on labor supply and effort

Unfortunately we do not possess data on individual performance or on the output produced by public employees. Therefore, we cannot evaluate whether the worsening of personnel selection and allocation processes due to corruption is also associated to lower quality of the public services being provided to the local community. However, through the labor force survey we can – at least partially – observe individual inputs, as measured by number of hours worked and a measure of absenteeism. As for the latter, we measure it by an indicator that equals 1 when the employee is on sick leave or absent for study and family reasons, as well as other causes which we deem to be more easily manipulated by the shirking worker (compare to more "exogenous" absence such as non-business days, compulsory maternity leaves, vacations, feast days, etc.).

In Table 3.13 we estimate an equation similar to (3.2) with (the logarithm of) number of hours worked and absence rate as dependent variables. When considering labor supply, we find that the coefficient associated to the interaction term between the public sector

[20] Among past foreign dominations, corruption is positively correlated with the Norman domination and negatively correlated, albeit to a lesser extent, with other spells of foreign domination, except for the Angevine and Swabian ones.

dummy and the measure of corruption is negative, suggesting that in areas with more corruption the number of hours worked by public employees relative to those in the manufacturing sector decreases. The coefficient for the subsample of managers and high-level professionals is twice in size as that for all non-manual workers. When considering absenteeism as a measure of individual effort, we find that the difference in the absence rate between the manufacturing and the public sector is indistinguishable for the entire sample but becomes statistically significant, albeit weakly so, when considering only employees at the top of the occupational ladder.

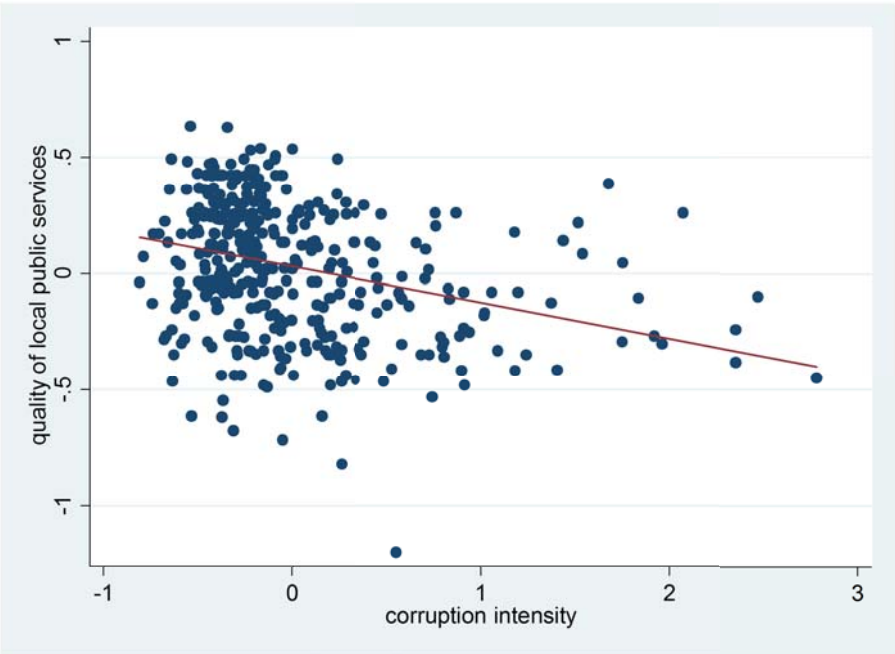


Figure 3.6: Corruption intensity (C^1) and quality of local public services (from Camussi and Mancini, 2016) across LLMs.

These results indicate that, besides worsening selection and allocation processes, corruption also leads to a relatively lower effort of public employees. We argue that these public employees' characteristics and behavior can (at least in part) explain the strong correlation existing between corruption and the quality of the administration's economic decisions and the effectiveness of the public services, as documented in earlier studies and as confirmed also by descriptive evidence reported in Figure 3.6.

3.5 Conclusions

Our analysis highlights the distortionary effect of corruption (as measured by local-level statistical indicators based on the number of reported crimes, on citizens' trust in local public institutions, on perception of administrations' integrity and on the quality of public expenditure) on the patterns of selection and allocation of public sector employees. Because of the nature of the tasks assigned to and areas of activity spanned by public agencies, public employees are more educated with respect to their counterparts in the private sector. This gap is, however, thinner where corruption indicators are higher, and the education bias induced by corruption is particularly strong for professions at the top of the occupational hierarchy. Similar evidence is found if we consider further dimensions of human capital such as grades obtained at school. Besides affecting selection, corruption contributes to deviating the education-based matching between workers and jobs: where corruption indicators are higher, public employees are relatively more likely to be assigned to tasks which are, on average, undertaken by more qualified personnel.

The comparative analysis of our results thus suggests that higher levels of corruption are associated with a poorer capacity of the public sector to select and allocate workers. Hence – if one believes that the workforce's human capital is conducive of better decision making – where corruption is high, the public administration will tend to adopt socially inefficient decisions and to remunerate individual less in terms of schooling ability than of other (unobserved) ability traits such as soft skills, relational capital or craftiness.

The eradication of corruption or, at least, the dampening of its implications have long been a major objective of governmental effort. Actions taken by governmental authorities usually rest on ex-post, repressive measures, which are sometimes accompanied by ex-ante, preventive provisions. The latter often take the form of a requirement for individual agencies to implement "in-house" anti-corruptive programs under governmental supervision. In light of the evidence presented in this paper, one may suspect that the administrations' ability to exert anti-corruptive self-monitoring might be hindered by corruption itself. Indeed, existing levels of crime in the environment may have contributed to the selection of a workforce which will, in general, be more likely to be misallocated as well as less prone to take up action against corruption if called to do so. The risk is that self-regulation aimed at overcoming corruption may work well only where corruption is already rare and fare poorly where it is more intense. Hence our results suggest caution against over-estimating the additional benefits of ex-ante, decentralized provisions.

Table 3.12: Robustness: IV estimates

Dependent variable:	Employed in the public sector			
Instrumental variable:	Past public dependence		Past dominations	
Professional area:	All	Managers/ professionals	All	Managers/ professionals
Years of schooling	0.016***	0.027***	0.016***	0.026***
	(0.001)	(0.002)	(0.001)	(0.002)
Years of schooling × C^1	-0.041***	-0.071***	-0.022***	-0.035***
	(0.006)	(0.012)	(0.003)	(0.006)
LLM × professional area FEs	YES	YES	YES	YES
F-stat of excluded instruments	119.2	35.1	22.8	7.0
R-squared	0.257	0.152	0.267	0.209
# observations	*397,060*	*99,663*	*397,060*	*99,663*
Dependent variable:	Under-education			
Public sector × C^1	0.104***	0.040	0.088***	0.072***
	(0.024)	(0.039)	(0.014)	(0.026)
LLM FEs	YES	YES	YES	YES
Sector of activity FEs	YES	YES	YES	YES
F-stat of excluded instruments	36.8	34.9	6.9	6.6
R-squared	0.044	0.059	0.044	0.059
Dependent variable:	Over-education			
Public sector × C^1	0.009	-0.015	0.005	-0.022
	(0.017)	(0.028)	(0.010)	(0.016)
LLM FEs	YES	YES	YES	YES
Sector of activity FEs	YES	YES	YES	YES
F-stat of excluded instruments	36.8	34.9	6.9	6.6
R-squared	0.043	0.038	0.043	0.038
# observations	*397,064*	*99,663*	*397,064*	*99,663*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variables are the following: employed in the public sector is equal to 1 for public sector employees and to 0 for manufacturing sector employees; under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (ISCO occupational at 3 digits). Years of schooling are those corresponding to the highest education attainment of the individual. Corruption is measured at the LLM level and we consider reported crimes net of judicial efficiency (C^1) instrumented with past public sector dependence (first two columns) and with length of the different past dominations (last two columns). Other controls include fixed effects for gender and age cohort.

Table 3.13: Labor supply and effort

Dependent variable:	Hours worked		Absenteeism	
Professional area:	All	Managers/ professionals	All	Managers/ professionals
Public sector × C^1	-0.016***	-0.032***	-0.001	0.003*
	(0.004)	(0.010)	(0.001)	(0.002)
LLM FEs	YES	YES	YES	YES
Sector of activity FEs	YES	YES	YES	YES
R-squared	0.397	0.427	0.005	0.010
# observations	*351,609*	*86,949*	*397,064*	*99,663*

Standard errors are clustered at the LLM level (* p<0.1, ** p<0.05, *** p<0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variables are (the logarithm of) hours worked (first two columns) and an indicator of absenteeism (last two columns). Corruption is measured at the LLM level with C^1, i.e. reported crimes net of judicial efficiency. Other controls include fixed effects for gender and age cohort and, when the dependent variable is hours worked, an indicator for whether the individual works part-time.

Chapter 4

M&A negotiations with limited information: how do opaque firms buy and get bought?

Co-authored with Pierpaolo Battigalli, Carlo Chiarella and Stefano Gatti.

4.1 Introduction

The consequences of informational frictions on corporate activities have been documented in various contexts, from a firm's underpricing in initial public offerings (e.g. Beatty and Ritter, 1986) to its cost of capital (e.g. Easley and O'Hara, 2005) or discount in private equity negotiations (e.g. Hertzel and Smith, 1993). However, the relationship between asymmetric information and bidding behavior in mergers and acquisitions (M&As) needs further investigation.[1] A rich collection of anecdotal evidence suggests that information asymmetry between target and bidder firms indeed results in frictions in the market for corporate control. In many cases, bidders eventually regret their *ex post* overpaid acquisitions and, on several occasions, bid valuation by targets with limited information has been so contentious as to end up in court.

Deal success clearly hinges on how much is paid and how.[2] In this paper, we model the negotiation between a bidder and a target as a signaling game with private information

[1] Haleblian et al. (2009) provide a comprehensive review of empirical findings on managerial behavior related to M&As from research in management, economics, and finance.

[2] For a more detailed discussion see, for example, Cording et al. (2002), Antoniou et al. (2008), and Gillis (2009).

on both sides and we examine a sample of M&A bids by and for U.S. publicly listed firms over the period $1985-2014$ to study the relations between three variables: the method of payment, the bid (acquisition) premium, and firm opacity (a measure of how hard it is for third parties to evaluate the firm's true value based on publicly available information and thus a proxy of how limited the information available to a counterparty in a transaction is[3]). In doing so, we assess the strategic rationale of the observed bidding behavior and the efficiency of the market for corporate control.

In particular, with asymmetrically informed counterparties, we expect firms' individual opacity to be an important driver of the simultaneous determination of both the bid premium and the method of payment. Since the probability of bid success increases with the value of the offer, but is accompanied by overpayment costs, bidders face an evident trade-off between the likelihood of overpaying and that of missing potential synergistic opportunities if their bid is rejected. Both these costs depend on the extent to which counterparties are privately informed and, most importantly, vary with bid premiums and across methods of payment. We then hypothesize that the informational structure of the deal affects how much is offered and how.

The theoretical underpinnings for our research are provided by the models of M&A under asymmetric information of Hansen (1987), Stultz (1988), Fishman (1989), Eckbo et al. (1990), and Rhodes-Kropf and Viswanathan (2004). These models show, on the grounds of alternative motivations, how the presence of informational asymmetries may have a significant impact on bid characteristics. Our aim in this paper is to unify the different views from these analyses under a novel theoretical framework. To this end, we propose a stylized bargaining game with asymmetric information on both sides, according to which the characteristics of actual bids will depend on the intensity and interaction of the informational gaps between counterparties, which originate from target and bidder opacity.

Our investigation of the choice of the method of payment and the expected bid premium for different degrees of opacity of bidder and target firms contributes to the M&A literature along several dimensions. First, while definitely stylized, our model allows us to capture the main features of the strategic interaction between a privately informed bidder and a privately informed target, in which the former simultaneously chooses the method of payment and the amount to be paid. By so doing, we depart from many existing

[3] We consider opacity as a firm characteristic and, in particular, we deem a firm to be more opaque when its valuation depends to a greater extent on unobservables (e.g. R&D based companies). Then, the informational asymmetry between two firms in a transaction is determined by their individual opacity or, in other words, by how limited the information of each counterparty is.

studies - theoretical and empirical - on the determinants of bid premiums, which consider the method of payment to be predetermined with respect to the amount offered. This allows us to show how bid features may depend not only on observable characteristics of the involved parties (e.g. firm size, deal materiality, ...) but also on each party's beliefs about the opponent's true value and, therefore, on how firms reason about each other's characteristics and actions, when precise information about the former is not available. Moreover, even under simplifying assumptions and unlike most previous work, our model gives rise to different kinds of equilibria under different parametric configurations. We are thus able to highlight the double channel through which opacity affects bids characteristics: on the one hand, different levels of opacity are associated with different kinds of equilibria (i.e. with different links between bidder characteristics and submitted offers); on the other hand, within each kind of equilibrium, opacity impacts directly on the likelihood of a given method of payment being used and on the average bid premium.

Second, by testing our hypotheses jointly and directly, our empirical analysis departs from existing studies which have so far typically focused on either the bidder or target side and, regarding the method of payment, have mainly drawn indirect inferences based on cumulative abnormal returns upon the announcement of a bid.[4] Indeed, Moeller et al. (2007) find that if stock is used as the exchange currency, abnormal returns to bidders are negatively related to their own extent of private information; Officer et al. (2009) report higher announcement returns for bidders using stock to acquire targets that are difficult to value. To the best of our knowledge, the only paper that presents a direct and joint test of the implications of both target and bidder private information on the choice of the method of payment is by Chemmanur et al. (2009). Still, the hypotheses they test do not coincide with the testable predictions that result from our theoretical model, which also includes the simultaneous determination of the bid premium. Analogously, existing studies have typically explored the relation between the bid premium and asymmetric information from the bidder's perspective only.[5] Indeed, Cheng et al. (2016) and Chatterjee et al. (2012) test how the opacity of target firms affects bid premiums, drawing on theories of overpricing due to divergence of opinion (e.g. Chen et al., 2002; Diether et al., 2002;

[4]For example, Travlos (1987), Amihud et al. (1990), Brown and Ryngaert (1991), and Servaes (1991) report significantly lower returns for bidders using stock instead of cash around the announcement date. Similarly, Franks et al. (1988) reveal that targets' returns are higher if they are offered cash instead of stock.

[5]For example, Koeplin et al. (2005) document that private firms are acquired at an average $20\% - 30\%$ discount relative to acquisition multiples (earnings) of similar publicly traded firms. The authors argue that the discount may partly be risk compensation to the bidder for adversely selecting a potential 'lemon' target. However, for public targets, Fuller et al. (2002) and Officer (2007) find that lower prices are paid for targets whose stock is less liquid.

Miller, 1977). In this respect, we extend their analysis by also taking into consideration the opacity of the bidder.

Third, to the best of our knowledge, studies on M&As have so far only measured firm-specific opacity, the cross-sectional conditioning variable, on the basis of *ex ante* firm characteristics. For example, Chemmanur et al. (2009) employ the number of analysts following a firm, the dispersion of their earnings per share (EPS) forecasts, and their forecast errors, while Chatterjee et al. (2012) use the dispersion of analysts' EPS forecasts, the breadth of mutual fund ownership, and idiosyncratic volatility. In this respect, we further contribute to the M&A literature by proposing to capture firm-specific opacity from a firm's equity trading properties instead, forming an index on the basis of the first principal component of several proxies for adverse selection risk from the literature on market microstructure, as in Bharath et al. (2009). Market microstructure measures of information asymmetry, and of adverse selection risk in particular, are in fact designed to capture investors' perception of the informational advantage held by firm insiders. For the sake of our analysis, then, they provide us with a more direct representation of the informational gaps between counterparts in a transaction than ex-ante firm characteristics.

Our empirical analysis first documents that the opacity faced by the bidder in assessing the value of the target is a significant driver of the choice of the method of payment. Indeed, consistent with the predictions of our model, the likelihood of a stock bid increases with the opacity of the target. The latter is indeed positively correlated with the likelihood that the bidder offer stock independently of its own value, so to alleviate overpayment concerns. We do not find evidence on the use of cash bids as a signaling device to deter potential competitors' bids for more opaque targets, as do Chemmanur et al. (2009), but, in line with the same preemptive bidding rationale and with Chatterjee et al. (2012), we find instead target opacity to be associated with higher bid premiums, as predicted by our model. Our analysis then documents that the opacity of the bidder is related to lower premiums, consistent with the fact that bidding firms take advantage of targets' impaired ability to assess their value when conditions are such that bidders with different values do not use separate means of payments. Other results are discussed in greater detail in Section 3.

The rest of the paper is organized as follows. Section 2 describes the model for the choice of the method of payment and the bid premium and formulates testable hypotheses concerning the impact of firm opacity on observable bid characteristics. Section 3 introduces the sample, describes the methodology, the index of firm opacity and presents the results, comparing the empirical findings with the theoretical predictions. Section 4 concludes the paper and introduces potential developments for further research.

4.2 The Model

To guide the construction of the relevant empirical hypotheses, we model the negotiations between a bidder and a target in the context of a two-stage Bayesian game. In particular, we consider a framework in which a bidder takes advantage of synergistic opportunities upon the acquisition of a target. The informational structure of the interaction is characterized by asymmetric information about the true, unobservable value of the counterpart and, consequently, the potential benefits from the transaction, with each firm being privately informed about its own stand-alone value. For both bidder and target, market values may not reflect the true value of the firm. The extent of uncertainty outsiders encounter in assessing the other firm's value, i.e. opacity, captures partially unobservable, firm-specific characteristics.

Wealth-maximizing counterparties negotiate, comparing their expected wealth gain conditional on alternative methods of payment and different bid premiums on the basis of the information they possess. We do not consider bids that combine stock and cash payment. A target firm satisfies its incentive constraint by accepting only bids in excess of its true value. When cash is offered, the value of the offer is independent of the true value of the target and the bidder bears the entire cost of overpayment. The probability of bid success and expected overpayment costs increase in the value of the bid (and the premium) and depend only on the value of the target.

On the other hand, in the case of a stock bid, the target is offered shares of the combined firm at some exchange ratio and needs to judge the value of the bid (and the premium) on the basis of its limited information. The terms of the offer are contingent and overpayment costs accruing to the bidder are reduced, since the target eventually shares gains and losses from the deal. However, the probability of bid success and expected overpayment costs depend not only on the value of the target, as in the case of a cash offer, but also on the target's assessment of the value of the combined firm. Stock offers then provide additional flexibility to satisfy the incentive constraints imposed by the presence of private information, but also entail additional informational costs.

4.2.1 Ingredients

We model the interactive situation between a generic bidder and its[6] target by a two-stage Bayesian game with two-sided asymmetric information. Such a game can be represented

[6]Both players are firms and will therefore be referred to by the neutral pronoun 'it'. Each firm acts as a single decision maker and we shall then talk, for instance, about 'a firm's actions', 'a firm's behavior', ...

as

$$G = \left\langle \{B,T\}, M, \{Y,N\}, \left(\Theta_i, p^i, u_i\right)_{i\in\{B,T\}} \right\rangle . \tag{4.1}$$

$\{B,T\}$ is the set of players, containing one bidder and one target. The role of player B in this model is to submit an *offer* directed to player T. An offer is an agreement specifying

- a method of payment, cash (C) or stock (S);
- if the method of payment is cash, an amount $c \in \mathbb{R}_+$ to be transferred from the bidder to the target;
- if the method of payment is stock, a participation share (or fraction) $f \in [0,1]$ to be awarded by the bidder to the target.

Thus the set of actions available to player B is

$$M := (\{C\} \times \mathbb{R}_+) \cup (\{S\} \times [0,1]) . \tag{4.2}$$

An action $m \in M$ available to the bidder is also called *message* (whence the choice of the letters m and M). The role of player T in this model is to either accept or reject the bidder's offer: its action set is $\{Y,N\}$, where Y denotes acceptance and N denotes rejection of the bidder's offer. An action $r \in \{Y,N\}$ available to the bidder is also called *response* (whence the choice of r to denote it).

Both firms have characteristics which are known to themselves, but unknown to the counterpart; that is, they possess private information. For each firm, each distinct set of characteristics constitutes an information type. In our model, the only relevant unknown characteristic of an agent is its *stand-alone value*, a positive real number summarizing the firm's value before the interaction takes place. We assume that, for each $i \in \{B,T\}$, the set of possible asset values of i is $\Theta_i := \left\{\underline{\theta}_i, \overline{\theta}_i\right\}$ where $0 < \underline{\theta}_i < \overline{\theta}_i$. For simplicity, we call these the 'low' and 'high' value (type) of agent i and we denote the difference among them by $V_i := \overline{\theta}_i - \underline{\theta}_i$. For all $i \in \{B,T\}$, $p^i \in (0,1)$ is agent i's belief that the opponent's value is $\overline{\theta}_j$. Such beliefs are commonly known by the players.[7]

[7] While beliefs are always an element of the players' subjectivity from the modeling point of view, we interpret the values in Θ_i as observable elements which every potential opponent would agree on and, conversely, we interpret p^i as the output of i's processing of available information: if i's role were to be played by another agent, the latter might hold a different belief. We will, therefore, sometimes call p^i the "subjective belief" of player i. Still, in our model, the information gathered by i and i's way to process it are known to i's opponent, j, and i knows that j knows, and so on, which makes p^i common knowledge.

For each $i \in \{B, T\}$, the (net) payoff function $u_i : M \times \{Y, N\} \times \Theta_B \times \Theta_T \to \mathbb{R}$ associates with every pair of action profiles and stand-alone values the monetary payoff obtained by i in excess of θ_i. Both firms are assumed to be risk-neutral.

The structure of the game is as follows. The bidder moves first and proposes an offer $m \in M$. The target observes the offer proposed by B and either accepts it (Y) or rejects it (N). If B's offer is accepted, the merged firm BT is created. We assume that the value of the merged firm BT, conditional on the stand-alone values of the involved parties, is given by $\theta_B + w(\theta_T)$, where, for all $\theta_T \in \Theta_T$, $w(\theta_T) > \theta_T$. Upon acquisition by the bidder, the value of the target's assets increases. This transformation in values is called *synergy* and we use the same name to denote the intensity of such increase, namely

$$\Delta(\theta_T) := w(\theta_T) - \theta_T \,. \tag{4.3}$$

Such synergy is commonly known. For simplicity, we let $W := w(\overline{\theta}_T) - w(\underline{\theta}_T)$. Thus W is the difference in post-merger values between a high- and a low-type target.

We make the following assumptions on our parameters:

(A1) $W > 0$: synergies preserve the ordering of types so that, *ceteris paribus*, the bidder finds it more convenient to acquire a high-type target than a low-type target;

(A2) $\frac{\overline{\theta}_T}{\underline{\theta}_T} \geq \frac{\overline{\theta}_B + w(\overline{\theta}_T)}{\underline{\theta}_B + w(\underline{\theta}_T)}$: the ratio among the two possible stand-alone values of the target is larger than the ratio among the values resulting from a 'high-high' and a 'low-low' merge.

While assumption (A1) is easy to interpret, assumption (A2) deserves further consideration. Although its function is mainly technical, (A2) simultaneously requires that the variability in stand-alone values is larger for the target than for the bidder (since it implies $\frac{\overline{\theta}_T}{\underline{\theta}_T} \geq \frac{\overline{\theta}_B}{\underline{\theta}_B}$) and yet that such variability in the target's standalone value has sufficiently small impact on the variability of the merged firm's value (the numerator and denominator of the fraction on the right-hand side of the expression are, respectively, the largest and smallest value of a merged firm). In other words, (A2) can be interpreted at one time as an assessment of a more difficult evaluation of the target's value than of the bidder's by a non-informed observer and as a requirement imposing that the post-merger value of low-type targets be relatively large compared to that of high-type targets (a form of 'diminishing returns to scale'). This can be seen by considering the two 'extreme' cases. If one assumes that, for each $i \in \{B, T\}$ and for some $\lambda \in \mathbb{R}_{++}$, $\overline{\theta}_i = (1 + \lambda)\, \underline{\theta}_i$, assumption (A2) requires that $w(\overline{\theta}_T) \leq (1 + \lambda)\, w(\underline{\theta}_T)$; if, on the other hand, one assumes that for some $\delta \in \mathbb{R}_{++}$, $w(\overline{\theta}_T) = (1 + \delta)\, w(\underline{\theta}_T)$, (A2) imposes that $\frac{\overline{\theta}_T}{\underline{\theta}_T} \geq \frac{\overline{\theta}_B}{\underline{\theta}_B}$.

If the bidder proposes a cash offer (C, c) and the target accepts it, then B obtains the value of the merged firm BT and T obtains the amount offered by B in cash:

$$u_B\left((C,c), Y, \theta_B, \theta_T\right) = w\left(\theta_T\right) - c\,, \tag{4.4}$$

$$u_T\left((C,c), Y, \theta_B, \theta_T\right) = c - \theta_T\,. \tag{4.5}$$

If the bidder proposes a stock offer and the target accepts it, then the two players share the value of the merged firm BT, with a fraction f going to T and the remaining going to B:

$$u_B\left((S,f), Y, \theta_B, \theta_T\right) = (1-f)w\left(\theta_T\right) - f\,\theta_B\,, \tag{4.6}$$

$$u_T\left((S,f), Y, \theta_B, \theta_T\right) = f\left(\theta_B + w\left(\theta_T\right)\right) - \theta_T\,. \tag{4.7}$$

If the bidder proposes an offer $m \in M$ which the target rejects, then both firms retain their stand-alone value, i.e. for each $i \in \{B, T\}$ and every offer $m \in M$, $u_i(m, N, \theta_B, \theta_T) = 0$.

4.2.2 Opacity and bid premium

We now define the two key variables of our analysis. We call *opacity* of firm i the number

$$\omega_i := p^j\left(1 - p^j\right)\,. \tag{4.8}$$

Calling σ_i^2 the variance attributed by j to i's post-merger value, the previous definitions state that ω_i is the ratio between this variance and the squared range of the post-merger value. Through this normalization, we make opacity depend exclusively on the subjective uncertainty of the counterpart and not on the objective variation in possible stand-alone values. Opacity thus quantifies how hard it is for the opponent to give a precise estimate of i's future value before the interaction takes place, by measuring the informativeness of firm j's beliefs.

We call *(relative) bid premium* the ratio between the difference in the target's payoff and its stand-alone value (the *absolute bid premium*) and the stand-alone value itself: for every action sequence $(m, r) \in M \times \{Y, N\}$ and every pair of stand-alone values $(\theta_B, \theta_T) \in \Theta_B \times \Theta_T$, the bid premium is

$$\psi(m, r, \theta_B, \theta_T) := \frac{u_T(m, r, \theta_B, \theta_T)}{\theta_T}\,. \tag{4.9}$$

The bid premium represents the relative profitability of the acquisition for the current management of the target (its value is 0 in case of a rejection by the target). Because the

M&A interaction results in a transfer of wealth from the bidder to the target in exchange for ownership, it is fair to see the bid premium as a price paid by the bidder to the target, although it does not coincide with the amount of money or the money-value of shares actually transferred. Notice, from (4.4),(4.5),(4.6) and (4.7), that, for all $m \in M$,

$$u_B\left(m, Y, \theta_B, \theta_T\right) = \Delta(\theta_T) - \theta_T\, \psi(m, Y, \theta_B, \theta_T) \tag{4.10}$$

so that, conditional on accepted offers, expected payoff maximization and expected bid premium minimization yield the same outcome.

4.2.3 Equilibrium concept

The equilibrium concept we use in this paper is that of *Perfect Bayesian Equilibrium* (see, for instance, Fudenberg and Tirole, 1991).

Definition 1. *A Perfect Bayesian Equilibrium* (PBE) *for the game G in (4.1) is composed of a strategy* $\mathbf{m}^*$ *of the bidder and an* assessment *(belief-strategy pair)* $(\boldsymbol{\mu}^*, \mathbf{r}^*)$ *of the target, where*[8] $\mathbf{m}^* \in M^{\Theta_B}$, $\boldsymbol{\mu}^* \in \Delta(\Theta_B)^M$ *and* $\mathbf{r}^* \in \{Y, N\}^{\Theta_T \times M}$, *such that*

1. *for every* $\theta_B \in \Theta_B$,

$$\mathbf{m}^*\left(\theta_B\right) \in \arg\max_{m \in M} \mathbb{E}_{p^B}\left[u_B\left(m, \mathbf{r}^*\left(\theta_T, m\right), \theta_B, \theta_T\right)\right]\ ; \tag{4.11}$$

2. *for every* $m \in M$, $\boldsymbol{\mu}^*(m)$ *is obtained from* p^T *via Bayes' rule, whenever possible;*

3. *for every* $\theta_T \in \Theta_T$ *and* $m \in M$,

$$\mathbf{r}^*\left(\theta_T, m\right) \in \arg\max_{r \in \{Y,N\}} \mathbb{E}_{\boldsymbol{\mu}^*(m)}\left[u_T\left(m, r, \theta_B, \theta_T\right)\right]\ . \tag{4.12}$$

Part (1) of the definition states that, for each type θ_B, action $\mathbf{m}^*(\theta_B)$ maximizes the bidder's expected payoff, given the target's equilibrium response. Randomness in the bidder's payoff is due to the bidder's uncertainty about the target's stand-alone value θ_T. Part (3) of the definition similarly requires that, for each message m received by the target, for each type θ_T and given equilibrium beliefs $\boldsymbol{\mu}^*$, the action prescribed by the strategy $\mathbf{r}^*$ upon receiving message m maximizes the type-θ_T target's expected payoff, computed under the equilibrium posterior beliefs $\boldsymbol{\mu}^*(m)$. Randomness in the target's payoff is due to the target's uncertainty about the bidder's stand-alone value θ_B.

[8] If X and Y are sets, Y^X denotes the set of functions from X to Y.

Part (2) of the definition requires that the target's posterior beliefs be obtained via Bayesian updating of its prior belief p^T. Whenever the message m sent by the bidder is consistent with the bidder's equilibrium strategy $\mathbf{m}^*$, this implies[9]

$$\boldsymbol{\mu}^* \left(\overline{\theta}_B \mid m\right) = \frac{p^T \mathbf{1}_{\mathbf{m}^*\left(\overline{\theta}_B\right)}(m)}{p^T \mathbf{1}_{\mathbf{m}^*\left(\overline{\theta}_B\right)}(m) + \left(1 - p^T\right) \mathbf{1}_{\mathbf{m}^*\left(\underline{\theta}_B\right)}(m)} . \tag{4.13}$$

If a message is received that is *not* consistent with the bidder's equilibrium strategy, the target can adopt any posterior belief.

From now on, to ease notation we shall write the maximand in (4.11), given a PBE $(\mathbf{m}^*, (\boldsymbol{\mu}^*, \mathbf{r}^*))$, as

$$U^*_{\theta_B}(m) := \mathbb{E}_{p^B}\left[u_B\left(m, \mathbf{r}^*\left(\theta_T, m\right), \theta_B, \theta_T\right)\right] . \tag{4.14}$$

Some PBE may be justified by assuming that the target, upon receiving an out-of-equilibrium offer $m \notin \mathbf{m}^*(\Theta_B)$, form posterior beliefs that are incompatible with the target assigning to the bidder the highest degree of strategic sophistication consistent with observed actions (e.g. by assuming that offer m comes with positive probability from a type $\theta_B \in \Theta_B$ for whom m is dominated). To deal with this, we shall make use of two refinements of PBE proposed by Cho and Kreps (1987) and known as the *Equilibrium Dominance Test* (EDT) and the *Intuitive Criterion Test* (ICT). EDT compels us to exclude equilibria in which posterior beliefs assign positive probability to the observed message having been sent by a type of bidder for which that message is unequivocally worse than its equilibrium action. ICT excludes equilibria in which the equilibrium message is optimal for some type of bidder only insofar as the latter expects the target to form beliefs which are incompatible with EDT.[10] While the application of EDT reduces the set of PBEs in our model by excluding some equilibrium assessments, it turns out that all equilibrium outcomes survive the ICT.

4.2.4 Cash offers

We first analyze cash offers, i.e. messages $m \in \{C\} \times \mathbb{R}_+$. Notice that the payoff gain that the target obtains when accepting a cash offer does not depend on the bidder's type. Hence the target's response to a cash offer will not depend on its beliefs about the bidder. It follows that the bidder does not value cash offers as signals of its own type.

[9] Here and in what follows, for all $m \in M$ and $\theta_B \in \Theta_B$, we write $\mu(\theta_B \mid m)$ instead of $\mu(m)(\{\theta_B\})$. Moreover, notice that, since we restrict our attention to pure-strategy equilibria, for each $m \in \mathbf{m}^*(\Theta_B)$, either $\boldsymbol{\mu}^*(\overline{\theta}_B \mid m) = p^T$ (pooling equilibrium) or $\boldsymbol{\mu}^*(\overline{\theta}_B \mid m) \in \{0, 1\}$ (separating equilibrium).

[10] A more thorough treatment of EDT and ICT for our model, based on Battigalli and Siniscalchi (2002), is given in Appendix 4.5.1.

Assuming that, when indifferent, a target always accepts the offer, it is easy to see that a cash offer (C, c) is accepted by the target if and only if $c \geq \theta_T$. Thus if $c \geq \overline{\theta}_T$ every type of target will accept the offer; if $\underline{\theta}_T \leq c < \overline{\theta}_T$, only the low-type target will accept the offer; finally, if $c < \underline{\theta}_T$ both types of target will turn the offer down.

As $p^B \in (0, 1)$, cash offers with $c \in (\underline{\theta}_T, \overline{\theta}_T)$ are dominated by $(C, \underline{\theta}_T)$ and cash offers with $c > \overline{\theta}_T$ are dominated by $(C, \overline{\theta}_T)$. Assumption (A1) is then enough to guarantee that, in every PBE where a cash offer is proposed by some type of bidder, such offer is either $(C, \overline{\theta}_T)$ (the *"high cash offer"*) or $(C, \underline{\theta}_T)$ (the *"low cash offer"*). For all $\theta_B \in \Theta_B$, the expected payoff obtained by a type-θ_B bidder with each cash offer is

$$\mathbb{E}_{p^B}\left[u_B\left((C, \underline{\theta}_T), \mathbf{r}^*(\theta_T, (C, \underline{\theta}_T)), \theta_B, \theta_T\right)\right] = (1 - p^B)\Delta(\underline{\theta}_T)\,, \tag{4.15}$$

$$\mathbb{E}_{p^B}\left[u_B\left((C, \overline{\theta}_T), Y, \theta_B, \theta_T\right)\right] = \mathbb{E}_{p^B}\left[w(\theta_T)\right] - \overline{\theta}_T\,, \tag{4.16}$$

where

$$\mathbf{r}^*(\theta_T, (C, \underline{\theta}_T)) = \begin{cases} Y & \text{if} \quad \theta_T = \underline{\theta}_T \\ N & \text{if} \quad \theta_T = \overline{\theta}_T \end{cases} \quad . \tag{4.17}$$

Then, given the target's response (4.17), the bidder will prefer the high cash offer over the low cash offer if and only if

$$p^B \geq \frac{V_T}{V_T + \Delta(\overline{\theta}_T)} =: \pi_C\,. \tag{4.18}$$

The number $\pi_C \in (0, 1)$ is called *cash offer threshold*. If the bidder is sufficiently optimistic about the target ($p^B \geq \pi_C$), then it will prefer the high cash offer to the low cash offer: it will have to pay more ($\overline{\theta}_T$), but it will have the target accept for sure and, in its opinion, the target is quite likely to be of the high type. Otherwise ($p^B < \pi_C$) the bidder will prefer the low cash offer, by which he acquires the target only if the latter is of the low type.

Notice that the cash offer threshold π_C does not depend on the bidder's type θ_B. The fact that the target's response to a cash offer does not depend on the bidder's type makes the bidder's choice between the two cash offers rely exclusively on the target's stand-alone value. This in turn implies that a cash offer cannot be an informative signal about the bidder's type for the target, as anticipated above. Therefore the target will not update its initial belief after observing an *equilibrium* cash offer.

4.2.5 Stock offers

Consider now stock offers, that is, messages $m \in \{S\} \times [0, 1]$. Let $\mu \in \Delta(\Theta_B)^M$ be the target's belief-updating rule. Upon receiving a stock offer (S, f), the target expects to

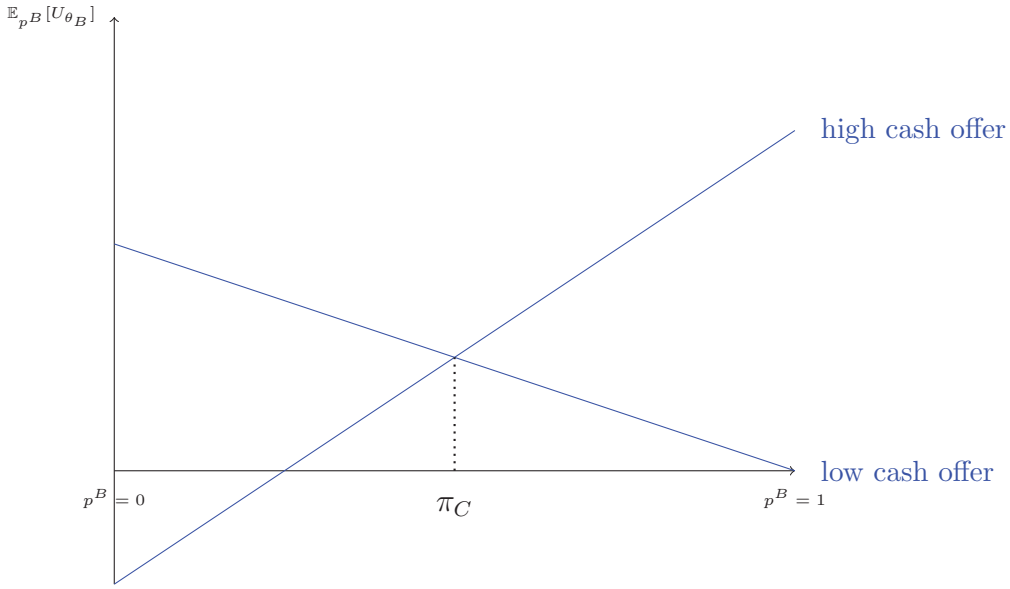


Figure 4.1: Expected payoffs from cash offers. The target's decision does not depend on the bidder's value, whence the common payoff for both types of bidder and the unique cash offer threshold π_C.

obtain (in case it accepts) a payoff equal to a fraction f of the post-merger value of the new firm BT,

$$\mathbb{E}_{p^T}\left[u_T\left((S,f),Y,\theta_B,\theta_T\right)\right] = f\left(\mathbb{E}_{\mu((S,f))}\left[\theta_B\right] + w(\theta_T)\right) - \theta_T\,. \tag{4.19}$$

Hence a type-θ_T target accepts the stock offer (S,f) if and only if

$$f \geq \frac{\theta_T}{\mathbb{E}_{\mu((S,f))}\left[\theta_B\right] + w(\theta_T)}\,. \tag{4.20}$$

Because of Assumption (A2), the right-hand side of (4.20) is larger for the type-$\overline{\theta}_T$ target, implying that a high-type target is choosier as far as stock offers are concerned, in the sense that whenever an offer (S,f) is accepted by a high-type target, it will also be accepted by a low-type target.

Unlike cash offers, the value of a stock offer to the target does depend on the latter's belief about the bidder's type, which in turn depends on the target's interpretation of the stock offer as a signal. One could not conclude *a priori* that if a certain type of target would accept the stock offer (S,f), then it would certainly also accept the stock offer (S,f') with $f' > f$. For some reason, the target might interpret a larger stock offer as a signal of there being a larger chance that the bidder has low type. In order to avoid

such situations, we require that, in and out of the equilibrium path, belief updating in our model satisfy the following assumption:

(A3) The function $f \mapsto \mu\left(\overline{\theta}_B \mid (S, f)\right)$ is continuous and non-decreasing.

Assumption (A3) refines the set of PBEs by requiring that a larger participation share in a stock offer does not induce the target to consider it less likely that the offer originated from a high-type bidder. In other words, the larger the fraction offered, the (weakly) larger the probability that the offer comes from a high-valued bidder.

Assumption (A3) guarantees that, given the updating rule μ, a type-θ_T target will accept any stock offer whose participation share exceeds a certain amount $f^{\theta_T}(\mu)$. Moreover, $f^{\overline{\theta}_T}(\mu) \geq f^{\underline{\theta}_T}(\mu)$: as expected, high-type targets are choosier, as it was the case with cash offers. For each $\theta_T \in \Theta_T$, the number $f^{\theta_T}(\mu)$ is determined as the fixed point of the map[11]

$$\Phi^{\theta_T}(f) := \frac{\theta_T}{\mathbb{E}_{\mu((S,f))}\left[\theta_B\right] + w(\theta_T)} \,. \tag{4.21}$$

We can then write the equilibrium optimal response of a type-θ_T target to stock offers as

$$\mathbf{r}^*(\theta_T, (S, f)) = \begin{cases} Y & \text{if} \quad f \geq f^{\theta_T}(\boldsymbol{\mu}^*) \\ N & \text{if} \quad f < f^{\theta_T}(\boldsymbol{\mu}^*) \end{cases} \,. \tag{4.22}$$

Given the equilibrium beliefs $\boldsymbol{\mu}^*$, the expected payoff of a type-θ_B bidder proposing a stock offer (S, f) with $f \geq f^{\overline{\theta}_T}(\boldsymbol{\mu}^*)$ is then

$$(1 - f)\mathbb{E}_{p^B}\left[w(\theta_T)\right] - f\,\theta_B \,, \tag{4.23}$$

as the target will always accept such an offer. If the proposed stock offer has $f^{\underline{\theta}_T}(\boldsymbol{\mu}^*) \leq f < f^{\overline{\theta}_T}(\boldsymbol{\mu}^*)$, it will be accepted only by a low-type target and therefore the bidder expects to obtain

$$(1 - p^B)\left[(1 - f)w(\underline{\theta}_T) - f\,\theta_B\right] \,. \tag{4.24}$$

Finally, if a stock offer is proposed with $f < f^{\underline{\theta}_T}(\boldsymbol{\mu}^*)$ then no target accepts it and the bidder retains its stand-alone value.

Whenever the offer is bound to be accepted by some type of target, the bidder's payoff is strictly decreasing in f and therefore, given the updating rule $\boldsymbol{\mu}^*$, the bidder will propose either one of the stock offers $(S, f^{\overline{\theta}_T}(\boldsymbol{\mu}^*))$ and $(S, f^{\underline{\theta}_T}(\boldsymbol{\mu}^*))$, respectively called the *"high"* and *"low stock offer"*.

[11] The existence and uniqueness of the fixed point is discussed in Appendix 4.5.2.

Thus the stock offer proposed by the bidder depends on the bidder's anticipation of the target's reaction, in terms of belief updating, to the bidder's message. In a PBE, the bidder correctly anticipates such reaction, holding a correct belief about the target's updating rule $\boldsymbol{\mu}^*$. For each type θ_B of bidder, a special case is what we call *"full recognition"* stock offers: these are the offers proposed by the type-θ_B bidder who wants to attract a certain type of target anticipating that, upon receiving the offer, the target will be almost sure that the bidder's type is θ_B. This must be the case, for instance, in any PBE in which a stock offer is proposed by one type of bidder but not by the other (i.e. in all 'separating' PBEs). When a stock offer which is just acceptable for type-θ_T targets is prescribed as the equilibrium message for a type-θ_B bidder in a PBE with full recognition, we abuse notation[12] and write

$$\mathbf{m}^*(\theta_B) = \left(S, f^{\theta_T}(\theta_B)\right) . \tag{4.25}$$

Thus $f^{\theta_T}(\theta_B)$ is the minimal participation share by which a type-θ_B bidder, which correctly anticipates to be recognized as such, can induce a type-θ_T target to accept its offer.

In Appendix 4.5.2 we prove the following

Lemma 2. *Under assumptions (A1)-(A3), if all stock offers lead to full recognition, then*

(a) for each $p^B \in (0,1)$, each type θ_B of bidder is indifferent between the low cash offer $(C, \underline{\theta}_T)$ and the low stock offer $(S, f^{\underline{\theta}_T}(\theta_B))$;

(b) for each $p^B \in (0,1)$, each type θ_B of bidder prefers the high stock offer $(S, f^{\overline{\theta}_T}(\theta_B))$ to the high cash offer $(C, \overline{\theta}_T)$;

(c) for each $\theta_B \in \Theta_B$, there exists a number $\pi_S(\theta_B) \in (0,1)$ such that a type-θ_B bidder prefers the high stock offer $(S, f^{\overline{\theta}_T}(\theta_B))$ to the low stock offer $(S, f^{\underline{\theta}_T}(\theta_B))$ if and only if $p^B \geq \pi_S(\theta_B)$;

(d) $\pi_S(\underline{\theta}_B) < \pi_S(\overline{\theta}_B) < \pi_C$.

According to Lemma (2), the expected payoff from a low full-recognition stock offer coincides with the payoff from the low cash offer, while the high full-recognition stock offer gives a larger payoff than the high cash offer. When advancing the low stock offer, the fully-recognized bidder promises to award the target with a share whose value equals $\underline{\theta}_T$,

[12]Here the participation share f implied by the stock offer is made to depend on the bidder's type rather than on its beliefs.

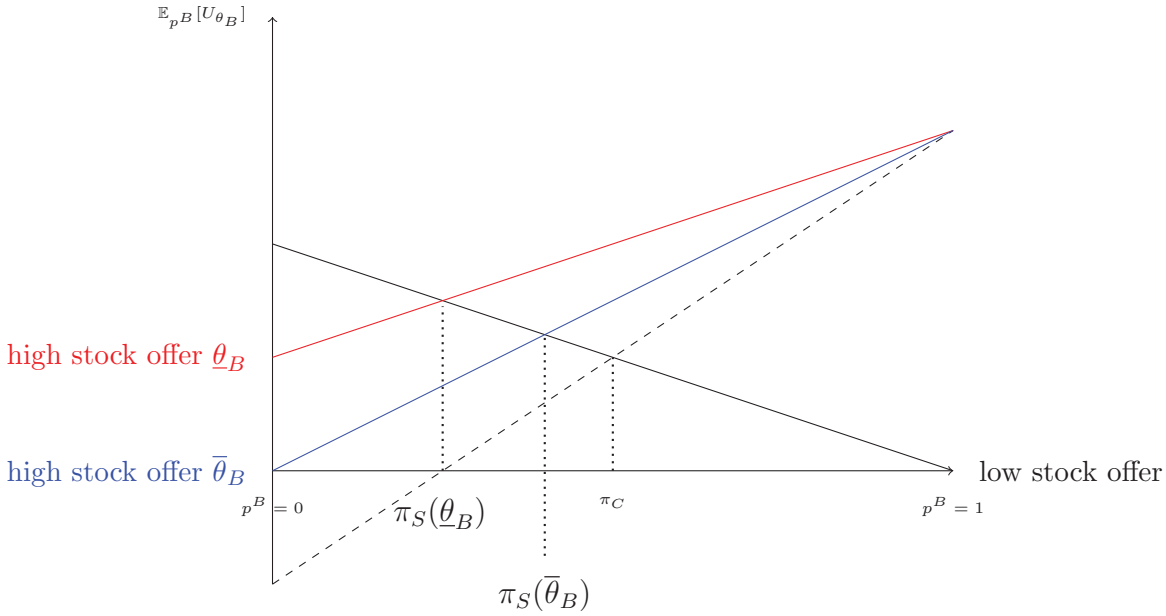


Figure 4.2: Expected payoffs from stock offers when both types of bidder are fully recognized. The blue line refer to the type-$\overline{\theta}_B$ bidder's high stock offer, red lines to the type-$\underline{\theta}_B$ bidder's high stock offer. The black line is the payoff from a low stock offer for both types of bidder. Dashed lines refer to the payoff accruing to each type of Bidder after a cash offer: only those from a high cash offer are visible in the diagram, as the low cash offer yields the same payoff as the low stock offer. As the payoff from stock offers is influenced by the bidder's value through the target's response as well, there are two distinct thresholds separating the regions where each stock offer is preferred to the other, one for the type-$\overline{\theta}_B$ bidder ($\pi_S(\overline{\theta}_B)$) and one for the type-$\underline{\theta}_B$ bidder ($\pi_S(\underline{\theta}_B)$).

conditional on $\underline{\theta}_T$ being the target's value. If the target is of the high type, the transaction does not occur. The low stock offer is therefore payoff-equivalent to the low cash offer when the bidder is fully recognized. In a high cash offer, the bidder has to promise $\overline{\theta}_T$ to the target, in order for it to accept, if it is of the high type. But the target will accept such offer independently of its type and, consequently, the bidder will be subject to an overpayment $\overline{\theta}_T - \underline{\theta}_T$ when it is matched with a low-type target. In a high stock offer, on the other hand, the bidder needs only convince the high-type target that the share it is being offered has value $\overline{\theta}_T$ *conditional on the target being of the high type.* The value of the share actually transferred to the target depends now on the target's type as well, with the consequence that the share obtained by a low-type target when accepting a high stock offer is valued more than $\underline{\theta}_T$ *but less than* $\overline{\theta}_T$. By making the actual value

of its offer contingent on the *ex ante* unknown value of the target, the bidder reduces its overpayment concern from $\overline{\theta}_T - \underline{\theta}_T$ to

$$\left(\frac{\theta_B + w(\underline{\theta}_T)}{\theta_B + w(\overline{\theta}_T)}\right)\overline{\theta}_T - \underline{\theta}_T\,. \tag{4.26}$$

It follows that, if full recognition were to occur for all types of bidder, no cash offer would ever be the bidder's unique best choice. The next paragraphs will clarify why this is not the case, justifying the observation of cash offers in the real world.

Figure 4.2 also shows that the gain from a stock offer is larger for the low-type bidder than for the high-type bidder. This is because the term in parentheses in (4.26) is increasing in θ_B. That is, $\underline{\theta}_B$'s disadvantage from having to offer a larger share of the merged firm is more than compensated by the fact that the share it will transfer to $\underline{\theta}_T$ is worth less, due to its own low value. The same reasoning clarifies why the lack of full recognition advantages the low-type bidder and penalizes the high-type bidder, as the former (latter) will need to offer a smaller (larger) share to induce the target into accepting.

4.2.6 Separating equilibria

A PBE strategy for the bidder, $\mathbf{m}^*$, must prescribe a message to each type of bidder. Because the bidder can have either of two stand-alone values, equilibria can be classified into two categories: those which prescribe the same message to each type of bidder (*pooling equilibria*) and those which don't (*separating equilibria*). Here we analyze the latter.

In a separating equilibrium, the two types of bidder will send different messages. Thus the (Bayesian) target will have no uncertainty about the bidder's type after observing the bidder's offer. In every separating equilibrium, for every $\theta_B \in \Theta_B$,

$$\boldsymbol{\mu}^*(\theta_B \mid \mathbf{m}^*(\theta_B)) = 1\,. \tag{4.27}$$

All such equilibria, then, display full recognition: upon observing the bidder's message, the target has no uncertainty about the bidder's stand-alone value. The following fact will prove particularly important.

Lemma 3. *Under assumptions (A1)-(A3),*

$$f^{\overline{\theta}_T}(\underline{\theta}_B) > f^{\overline{\theta}_T}(\overline{\theta}_B) \geq f^{\underline{\theta}_T}(\underline{\theta}_B) > f^{\underline{\theta}_T}(\overline{\theta}_B)\,. \tag{4.28}$$

Each type of bidder has to give up a larger share to convince the high-type target to accept the offer. Moreover, a low-type bidder will have to promise larger participation

than a high-type bidder to induce the same pool of targets to accept its offers, because the merged firm is worth less when the bidder has low type.

As our messages are two-dimensional (they consist of a method of payment and a real number), separating equilibrium candidates can be conveniently divided according to whether both types of bidder choose the same method of payment. As the optimal cash offer does not depend on the bidder's type, there can be no separating equilibrium in which both types of bidder propose a cash offer. Thus a separating equilibrium with common method of payment would need to be one in which both types of bidder propose a stock offer.

But every candidate separating equilibrium with both types of bidder offering stock falls apart as well. Because of the inequalities of Lemma (2-d), under full recognition it will never be the case that the high-type bidder prefers the high stock offer while the low-type bidder prefers the low stock offer. Then, by Lemma (3), when both types of bidder propose stock offers, the highest participation share is offered by the low-type bidder. The target would then deem it more likely that the bidder type is $\overline{\theta}_B$ after observing a lower stock offer, which contradicts assumption (A3). We make this reasoning rigorous in Lemma (11) of Appendix 4.5.2.

Inevitably, then, if a separating equilibrium exists, it must prescribe that the two types of bidder use different means of payments. If the high-type bidder were prescribed to offer stock, a stock offer would induce an almost-sure belief that its type is $\overline{\theta}_B$. In this case, though, the low-type bidder would have an incentive to deviate away from its optimal cash offer towards the high-type bidder's equilibrium offer. This excludes the possibility of such an equilibrium configuration, as we explain in Lemma (12) of Appendix 4.5.2.

We are finally left with the possibility that $\overline{\theta}_B$ proposes a cash offer, and that $\underline{\theta}_B$ proposes a stock offer. Notice in advance that, while a type-$\underline{\theta}_B$ bidder will always prefer a stock offer by which it is recognized as a type-$\overline{\theta}_B$ bidder to the optimal cash offer, the converse is not true: a type-$\overline{\theta}_B$ bidder will not always profit from abandoning cash in favor of a stock offer by which it could be mistaken for a type-$\underline{\theta}_B$ bidder. This is the reason why, under some conditions, a separating equilibrium in which $\overline{\theta}_B$ offers cash and $\underline{\theta}_B$ offers stock can exist.

Proposition 4. *If*

$$\frac{W\Delta(\overline{\theta}_T)}{V_B\left(V_T+\Delta(\overline{\theta}_T)\right)} \leq 1 \tag{4.29}$$

the game G in (4.1) admits, for all $p^B \in (0,1)$, a separating equilibrium in which $\mathbf{m}^*(\overline{\theta}_B)$ is a cash offer and $\mathbf{m}^*(\underline{\theta}_B)$ is a stock offer. If (4.29) does not hold, there are p^B_+, $p^B_- \in (0,1)$

such that $p_+^B > p_-^B$, $\pi_C \in \left(p_-^B, p_+^B\right)$ *and no separating equilibrium exists if* $p^B \in \left(p_-^B, p_+^B\right)$.

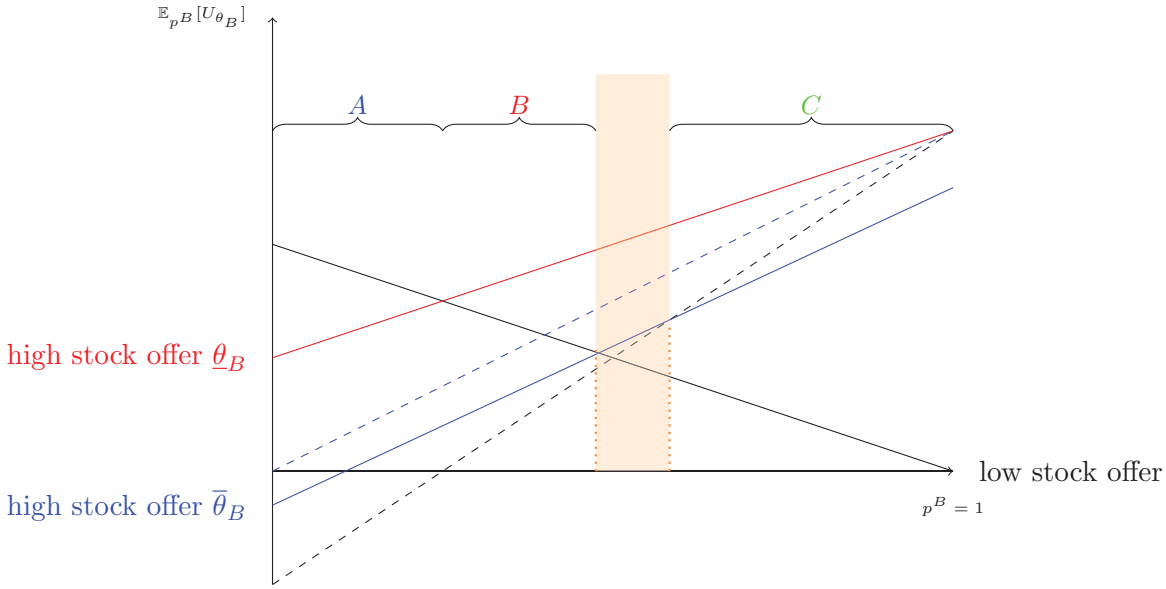


Figure 4.3: Separating equilibria. The shaded region highlights the interval where such separating equilibrium fails to exist, because the high-type bidder prefers to offer the low-type's high stock offer and be recognized as a low type (the payoff from this option is represented by the solid blue line), rather than proposing its best cash offer (dashed black line). The dashed blue line represents the high-type's bidder payoff from a high stock offer under full recognition.

When the low-type bidder is fully recognized, the optimal stock offer provides him with higher expected payoff than the optimal cash offer. Failure to support the separating equilibrium derives from a possible incentive for the high-type bidder to disguise itself as a low-type bidder: this forces the bidder to pay a larger participation share than it would in correspondence of *its* optimal stock offers but a smaller overall price than it needs to acquire the target by cash.

4.2.7 Pooling equilibria

We now turn to the analysis of *pooling equilibria*, which prescribe that $\mathbf{m}^*(\overline{\theta}_B) = \mathbf{m}^*(\underline{\theta}_B) = m^* \in M$. We consider first the case in which both types of bidder post the same cash offer. Notice first that this offer cannot be the high cash offer $(C, \overline{\theta}_T)$. Indeed, independently of the target posterior beliefs, the low-type bidder always prefers the high stock offer to

the high cash offer. By the same reasoning, one can exclude that the two types of bidder pool on the best cash offer whenever $p^B > \pi_S(\underline{\theta}_B)$.

The only candidate for an equilibrium message with pooling on a cash offer is, therefore, $\mathbf{m}^* = (C, \underline{\theta}_T)$. For all $p^B \in (0, \pi_S(\underline{\theta}_B)]$, the type-$\overline{\theta}_B$ bidder weakly prefers the low cash offer to its best stock offer. On the other hand, whenever there is a chance that it can be mistaken for a high-type bidder, the low-type bidder will prefer its best stock offer to the low cash offer. We then have to require that the low-type bidder is fully recognized at its low stock offer *and* that it does not want to deviate towards the high stock offer. The last condition requires that $\boldsymbol{\mu}^*(\overline{\theta}_B \mid (S, f^{\overline{\theta}_T}(\boldsymbol{\mu}^*)))$ be small enough. The equilibrium dominance criterion further forces such value to zero.

Proposition 5. *For every $p^B \leq \pi_S(\underline{\theta}_B)$ there exists an equilibrium in which $\mathbf{m}^*(\overline{\theta}_B) = \mathbf{m}^*(\underline{\theta}_B) = (C, \underline{\theta}_T)$.*

Let us finally focus on equilibria in which both types of bidder propose the same stock offer. First notice that, if the equilibrium message is $m^* = (S, f^*)$, then $\boldsymbol{\mu}^*(\overline{\theta}_B \mid (S, f^*)) = p^T$ by Bayes' rule. This excludes the possibility of both types of bidder pooling on a stock offer directed to the low-type target only: since $p^T \in (0, 1)$, the high-type bidder will not enjoy the benefit of full recognition and it will hence prefer to deviate towards the low cash offer. Thus at an equilibrium with pooling on stock offers, both types of bidder will propose the stock offer (S, f^*) with

$$f^* = \frac{\overline{\theta}_T}{\mathbb{E}_{p^T}(\theta_B) + w(\overline{\theta}_T)} \,. \tag{4.30}$$

Such an offer can be sustained as an equilibrium message only when the value of the best cash offer is sufficiently small and, at the same time, the prior target's belief that the bidder's type is $\overline{\theta}_B$ is large enough.

Proposition 6. *If*

$$p^T \geq 1 - \frac{W\Delta(\overline{\theta}_T)}{V_B\left(V_T + \Delta(\overline{\theta}_T)\right)} \tag{4.31}$$

there exist $p^B_{++}, p^B_{--} \in (0, 1)$ such that $p^B_{++} \geq p^B_{--}$ and the game G in (4.1) admits a PBE in which $\mathbf{m}^*(\overline{\theta}_B) = \mathbf{m}^*(\underline{\theta}_B) = (S, f^*)$, where f^* is given by (4.30), if and only if $p^B \in \left[p^B_{--}, p^B_{++}\right]$.

It turns out that equilibria with pooling on the stock offer complement separating equilibria whenever the latter fails to exist.

Corollary 7. *If (4.29) does* not *hold, then*

$$p^B_{--} \leq p^B_{-} < p^B_{+} \leq p^B_{++} \tag{4.32}$$

In other words, if p^B is such that G does not admit a separating equilibrium, then G admits a pooling equilibrium of the kind described in Proposition (6).

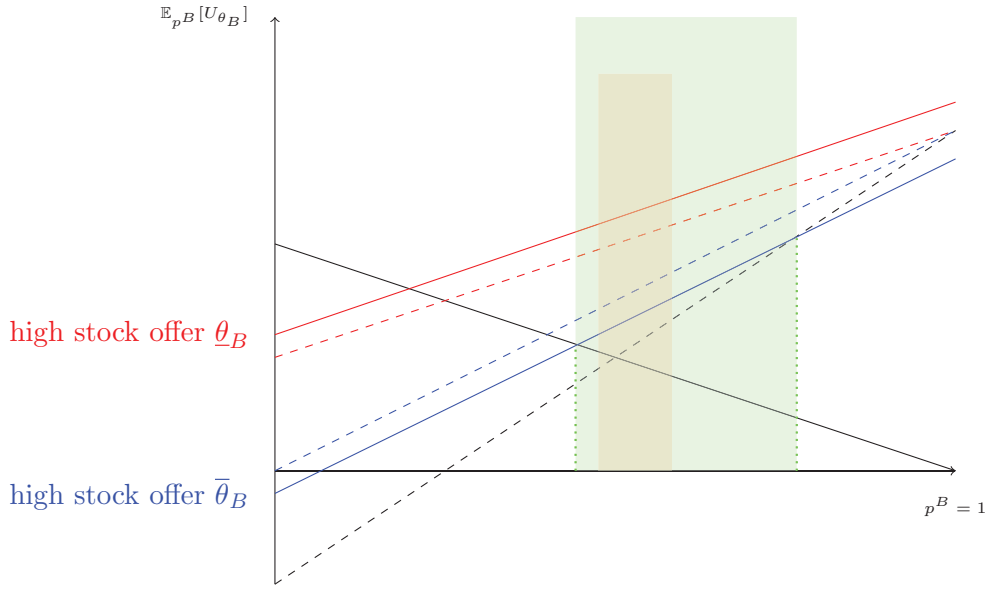


Figure 4.4: Stock pooling equilibrium. The solid blue and red lines represent, for each bidder, the payoff from a high stock offer upon receiving which the target does not update its prior belief p^T. The green shaded region represents the interval where the stock pooling equilibrium is sustained. Notice that this interval always includes the one where the separating equilibrium fails to exist (orange shaded region). The dashed blue and red lines are payoffs from the high stock offer under full recognition.

This result, paired with the previous Propositions, also guarantees that, for all parametric configurations satisfying assumptions (A1) and (A2), our model admits at least one equilibrium and at most two distinct equilibrium outcomes.

4.2.8 Opacity, means of payment and the bid premium

By now we know that three equilibrium configurations can arise in this game: a separating equilibrium in which the high-type bidder offers cash and the low-type bidder offer stock ;

a pooling equilibrium in which both types of bidder offer the low cash offer; and a pooling equilibrium in which both types of bidder offer the high cash offer. Our aim is now to inspect the correlation that these equilibrium predictions establish between, on the one side, opacity of either agents and, on the other, the probability of observing a stock offer and the average bid premium paid by the bidder.

Consider a game G as in (4.1) and let $e^* := (\mathbf{m}^*, (\boldsymbol{\mu}^*, \mathbf{r}^*))$ be a PBE of G. The probability that a transaction occurring under this equilibrium involves a stock offer is

$$\mathbb{P}^S(e^*) := p^T \mathbf{1}_{S\times[0,1]}(\mathbf{m}^*(\overline{\theta}_B)) + (1-p^T)\mathbf{1}_{S\times[0,1]}(\mathbf{m}^*(\underline{\theta}_B)) \,. \tag{4.33}$$

If e^* is a pooling equilibrium, one obviously has $\mathbb{P}^S(e^*) \in \{0,1\}$, the former case occurring when both types of bidder pool on a cash offer, the latter when both pool on a stock offer. If e^* is a separating equilibrium, the probability of observing a stock offer is the probability that the bidder's type is $\underline{\theta}_B$, which is $1-p^T$.

Recall now from (4.9) the definition of the (relative) bid premium. Given an equilibrium e^*, we define the average bid premium associated with e^* as

$$\psi(e^*) = \sum_{(\theta_B,\theta_T)\in\Theta_B\times\Theta_T} \mathbb{P}(\theta_B,\theta_T)\psi\left(\mathbf{m}^*(\theta_B), \mathbf{r}^*(\theta_T, \mathbf{m}^*(\theta_B)), \theta_B, \theta_T\right) \,. \tag{4.34}$$

Here $\mathbb{P}$ is the product probability derived from the priors of each player, which are considered independent.[13] The acceptance of a 'low' offer, be it cash or stock, generates no bid premium, since the type-$\underline{\theta}_T$ target, to which such offer is directed, receives an amount that makes it just indifferent between accepting and rejecting the offer. 'High' offers generate positive bid premiums only when the accepting target has type $\underline{\theta}_T$.

Thus, the average bid premium associated with a cash-pooling equilibrium is 0. Recall that in all equilibria with pooling on the stock offer, the two types of bidder direct their offer to both types of target. The average bid premium associated to a stock -pooling equilibrium e^{SP} will then be

$$\psi(e^{SP}) = (1-p^B)\left(\frac{\overline{\theta}_T}{\underline{\theta}_T}\cdot\frac{\mathbb{E}_{p^T}[\theta_B] + w(\underline{\theta}_T)}{\mathbb{E}_{p^T}[\theta_B] + w(\overline{\theta}_T)} - 1\right) \,. \tag{4.35}$$

Finding average premiums for separating equilibria requires more effort, because the kind of offers proposed depend on p^B. There are three cases: either both types of bidder direct their offer at the low-type target, and no bid premium is paid (*type-A separating equilibrium*: Figure 4.3, letter A); or the low-type bidder directs its stock offer to all

[13] So, for instance, $\mathbb{P}(\overline{\theta}_B, \overline{\theta}_T) = p^T p^B$ and, similarly, $\mathbb{P}(\overline{\theta}_B, \underline{\theta}_T) = p^T(1-p^B)$ and so on.

targets, while the high-type bidder acquires only low-type targets using cash (*type-B separating equilibrium*: Figure 4.3, letter B), in which case the average bid premium is

$$\psi(e^{S1}) = (1-p^B)(1-p^T)\left(\frac{\overline{\theta}_T}{\underline{\theta}_T}\cdot\frac{\underline{\theta}_B + w(\underline{\theta}_T)}{\underline{\theta}_B + w(\overline{\theta}_T)} - 1\right) ; \tag{4.36}$$

or, finally, both bidders aim at every type of target and realize different premiums with their offers of different kind (*type-C separating equilibrium*: Figure 4.3, letter C), yielding the average bid premium

$$\psi(e^{S2}) = (1-p^B)\left\{p^T\left(\frac{\overline{\theta}_T}{\underline{\theta}_T} - 1\right) + (1-p^T)\left(\frac{\overline{\theta}_T}{\underline{\theta}_T}\cdot\frac{\underline{\theta}_B + w(\underline{\theta}_T)}{\underline{\theta}_B + w(\overline{\theta}_T)} - 1\right)\right\} . \tag{4.37}$$

We simulate several games like G, letting the parametric configurations vary within the ranges that allow our assumptions (A1)-(A3) to hold.[14] For each equilibrium of each simulated game, we compute the probability of observing a stock offer at that equilibrium and the average bid premium. Suppose that the parameters of G are such that two equilibria, e^* and e°, arise. Keeping an agnostic stance on equilibrium selection, in such a case we shall associate to the parametric configuration under analysis a probability of observing a stock offer equal to the arithmetic mean of the two probabilities corresponding to the equilibria, that is $\frac{1}{2}\mathbb{P}^S(e^*)+\frac{1}{2}\mathbb{P}^S(e^\circ)$. Analogously, we shall associate to G an average bid premium equal to $\frac{1}{2}\psi(e^*) + \frac{1}{2}\psi(e^\circ)$. We the use these computations to inspect the overall correlations between bidder and target opacity and such observable variables. 4.1 summarizes the results of this simulation.

First, target opacity (ω_T) is positively correlated with the probability of observing a stock offer. Three forces stand behind this result. On the one hand, target opacity is larger when p^B takes on values near $\frac{1}{2}$, which makes the arising of a cash-pooling equilibrium less likely. On the other hand, the interval of beliefs p^B where a stock-pooling equilibrium can arise (which is bounded below by $\pi_S(\underline{\theta}_B)$) is relatively more likely to contain values around $\frac{1}{2}$ than values close to 1. This makes high values of opacity and the arising of stock pooling – which implies, by definition, the highest probability of observing a stock offer – positively correlated. This correlation is, finally, reinforced by the fact that intermediate values of p^B are associated with a larger likelihood of failure to obtain a separating equilibrium and, therefore, of having a unique, stock-pooling equilibrium[15]. Figures 4.5 and 4.6 show – for a sampled parametric configuration – how the four possible

[14] See the notes to Table 4.1 for details.

[15] For instance, in our simulations, stock pooling occurs in 28% of sampled parameter configurations such that $p^B \in [0.45, 0.55]$, and in more than 15% of those it is the unique equlibrium; while it arises in 20% of configurations such that $p^B \in [0.85, 0.95]$ and in only 6% of those it is the unique equilibrium.

Table 4.1: Simulations.

This table reports simulated correlations between the variables of interest in our model: target and bidder opacity, the probability of a stock offer being observes (MP) and the average bid premium (PRM). Correlation coefficients are estimated via 1,000 bootstrap samples of a set of 10,000 parameter vectors randomly drawn from $[0,1]^8$. The assumed distribution is uniform for all parameters except for p^B and p^T, which are assumed to be identically distributed according to the distribution reported under each panel. Each parameter vector which does not satisfy the model's assumptions is removed from the set: for each distributional assumption, the effective sample size is indicated below. The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Var. 1	Var. 2	Mean corr.	St. dev.	Mean corr.	St. dev.	Mean corr.	St. dev.
T-opacity	MP	0.1222***	0.017	0.1553***	0.018	0.0940***	0.017
B-opacity	MP	-0.0661***	0.020	-0.0851***	0.017	-0.0318	0.020
T-opacity	PRM	0.0631***	-0.010	0.1177***	0.016	0.0249***	0.009
B-opacity	PRM	-0.0001	0.013	-0.0161	0.017	-0.0003	0.011
Belief distr.		Beta(1,1)		Beta(.5,.5)		Beta(2,2)	
Eff. sample		2,794		3,032		2,914	

combinations of equilibria (separating and cash-pooling, separating only, separating and stock -pooling and stock -pooling only) are associated with p^B and progressively related with higher values of target opacity.

Bidder opacity (ω_B) is instead negatively correlated with the probability of observing a stock offer. The direction of this relationship, though, is less intuitive than it was for target opacity. In the context of separating equilibria, the probability of a stock offer decreases linearly with p^T. For parameter configurations at which a separating equilibrium exists (this does not depend on p^T), this implies no correlation between opacity and likelihood of a stock offer within separating equilibria. Larger values of p^T, though, are associated with the existence of stock pooling equilibria (see Figure 4.7, panels 3, 5 and 6) which naturally increase the *ex ante* likelihood of a stock offer. As shown in 4.8, the resulting correlation may vary in sign for given values of p^B. And, in particular, it may be positive when p^B is such that the threshold of Proposition (6) is small enough. Our simulations show that a concentration of the prior density of beliefs around central values leads to such correlation becoming insignificant.

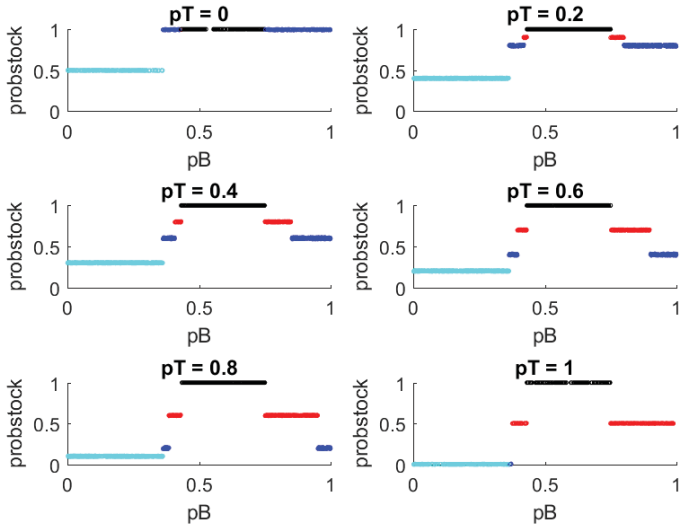


Figure 4.5: Probability of observing a stock offer over p^B. For the following figures parameters have been set to $\underline{\theta}_T = 1$, $\overline{\theta}_T = 3$, $\underline{\theta}_B = 2$, $\overline{\theta}_B = 5$, $w(\underline{\theta}_T) = 3$, $w(\overline{\theta}_T) = 5$. Black dots correspond to parameter configurations in which the unique equilibrium is a stock pooling equilibrium. Blue dots correspond to parameter configurations in which the unique equilibrium is a separating equilibrium. Red dots represent co-existence of a separating and a stock pooling equilibrium. Light blue dots represent co-existence of a separating and a cash pooling equilibrium.

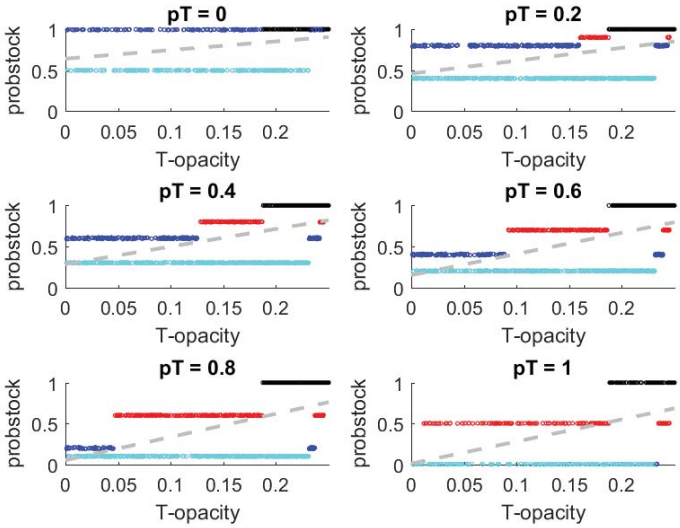


Figure 4.6: Probability of observing a stock offer over ω_T. The dashed grey line is the least-squares fitting line.

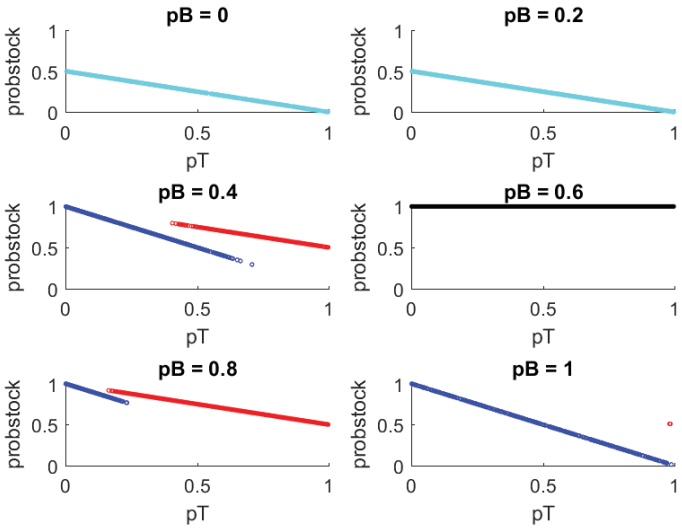


Figure 4.7: Probability of observing a stock offer over p^T.

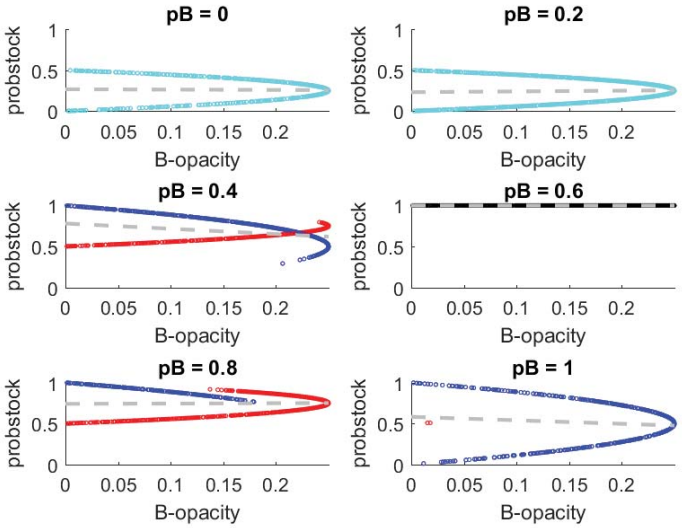


Figure 4.8: Probability of observing a stock offer over ω_B. The dashed grey line is the least-squares fitting line.

Target opacity turns out to be positively correlated with the average bid premium as well. Notice from 4.9 that, within each type of equilibrium, the average premium is decreasing in p^B. Moreover, one can verify that, when a separating equilibrium and a stock-pooling

equilibrium coexist, stock pooling yields a larger (respectively, smaller) premium in type-B (type-C) separating equilibria. The highest premiums are associated with the existence of stock pooling as the unique equilibrium and the lowest premiums are associated with either the low-low separating or with the high-high separating equilibrium as p^B nears 1. This generates (see Figure 4.10) a positive correlation between target opacity and bid premium similarly as to that between target opacity and the probability of stock being the chosen means of payment.

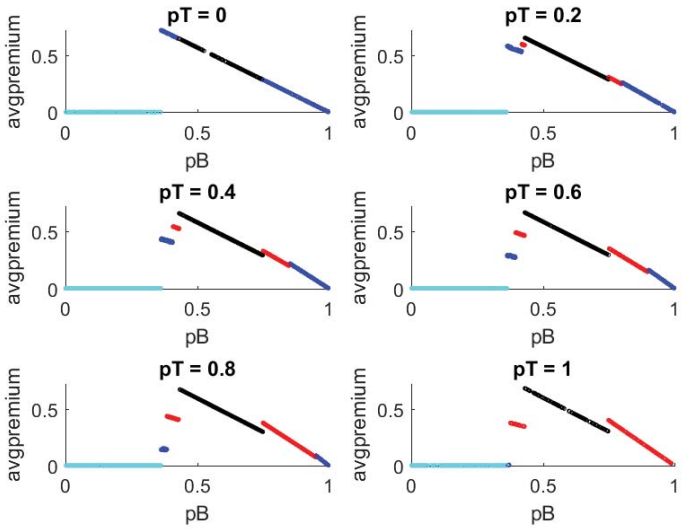


Figure 4.9: Average bid premium over p^B.

In our simulations, bidder opacity appears to be unrelated to the average premium. In the context of stock pooling equilibria, the premium increases in p^T, but less than linearly, so that bidder opacity is associated with an increase in premiums (Figures 4.15 and 4.16 in Apendix 4.5.4). In separating equilibria, premiums are linear in p^T (and, in particular, constant at zero in type-A separating equilibria, linearly decreasing in type-B separating equilibria and linearly increasing in type-C separating equilibria) so that, at parameter configurations at which a separating equilibrium exists, there is no correlation between bidder opacity and the bid premium (Figures 4.19 and 4.20 in Apendix 4.5.4). As Figures 4.11 and 4.12 illustrate, negative correlation arises only when parameters are such that a type-B separating equilibrium and a stock pooling equilibrium coexist (with the prevalence of stock pooling acting against a significant negative sign).

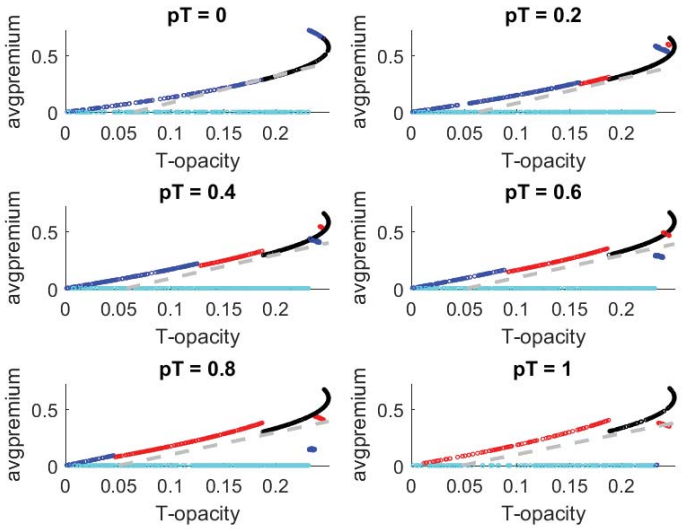


Figure 4.10: Average bid premium over ω_T. The dashed grey line is the least-squares fitting line.

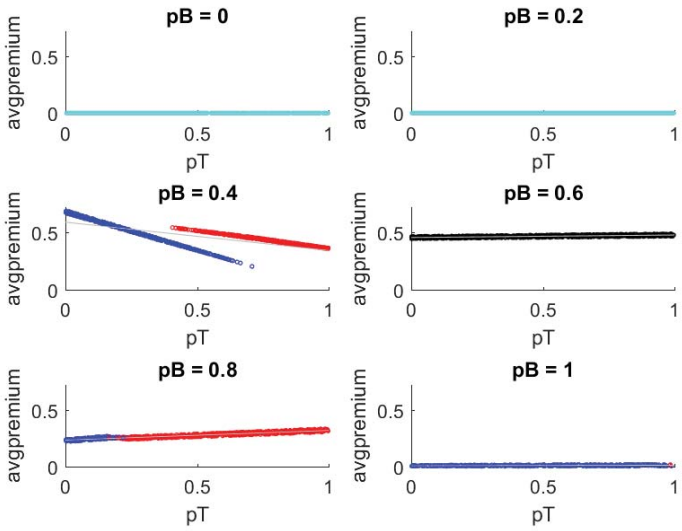


Figure 4.11: Average bid premium over p^T.

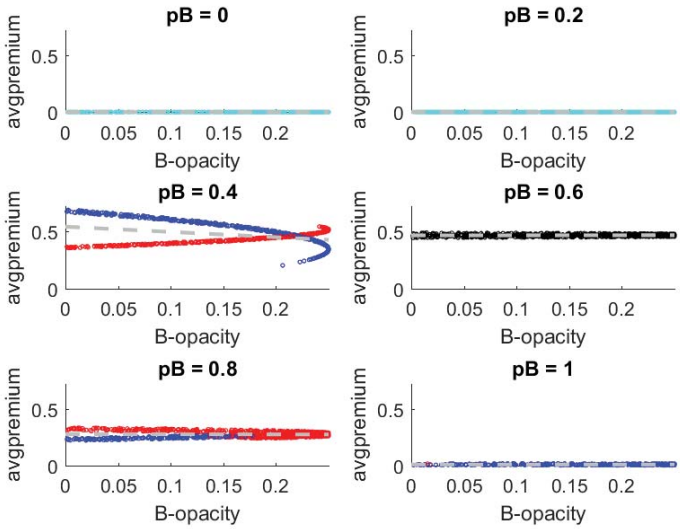


Figure 4.12: Average bid premium over ω_B. The dashed grey line is the least-squares fitting line.

4.2.9 A few hypotheses

Summing up the analysis in the previous section, we can collect the model's main suggestions into four testable predictions on how bidder and target opacity affect the method of payment chosen by the former and the bid premium to be paid by the bidder upon acceptance.

(i) the likelihood of a stock bid is positively correlated with the opacity of the target;

(ii) the likelihood of a stock bid is (possibly weakly) negatively correlated with the opacity of the bidder;

(iii) the bid premium is positively correlated with the opacity of the target;

(iv) the bid premium in uncorrelated (or, possibly, weakly negatively correlated) with the opacity of the bidder.

As we have pointed out before, predictions involving target opacity are not affected by assumptions used to perform simulations, while predictions involving bidder opacity may change if beliefs are not identically and/or not symmetrically distributed.

4.3 Empirical analysis

Our empirical analysis focuses on bid premiums in regard to how much, as a percentage, is offered for the acquisition of a target in excess of its stand-alone market valuation and on the qualitative dimension of the choice of the method of payment concerning the type of consideration used in the transaction among either cash or stock. For consistency with our model of bidding behavior under asymmetric information, other forms of payment, including mixed cash and stock, are excluded from our analysis. Mixed forms of payment, in particular, cannot be considered a distinct category for the sake of our analysis as there is a great heterogeneity among them depending on the different proportions of cash and stock they involve.

4.3.1 Data

Data on M&A announcements, as reported on SDC Thomson One Banker, are collected from 1985 to 2014. Both completed and withdrawn bids are considered.[16] We include in the sample only bids in which both the target and bidder firms are U.S. publicly listed non-financial firms.[17] We limit our sample to bids classified as mergers, acquisitions, or acquisitions of a majority interest. We then exclude buybacks, exchange offers, recapitalizations or acquisitions of a partial or remaining interest. These restrictive requirements are expected to result in a sample of transactions for which asymmetric information is a potentially important concern, since a bidder who did not previously own a majority interest in the target is indeed seeking to obtain a majority interest through the transaction.[18] We consider only transactions whose reported value is in excess of $50 million, adjusted for inflation and expressed as 2014 equivalents. We exclude deals for which consideration is not reported as either cash or stock and for which the combined amount of cash and stock accounts for less than 95% of the transaction value. Finally we exclude transactions in which the bidder is a financial sponsor or the target is a subsidiary, a joint venture or government owned.

Our final sample consists of 3141 bids. Still, the number of actual bids used for our analysis is constrained by the availability of the relevant data on bidder and target opacity and

[16] We believe that firm opacity can result in different distributions of withdrawals across methods of payment. The inclusion of both successful and withdrawn bids then reduces potential concerns of selection bias.

[17] Firms whose main business activity is classified within Standard Industrial Classification (SIC) codes 6000-6999 are considered financial firms.

[18] These requirements are in line with the work of Chemmanur et al. (2009).

other firm-level characteristics. Data on M&A bids and deal characteristics from SDC Thomson One Banker are in fact complemented with firm-level stock market data and financial data from the Center for Research in Security Prices (CRSP) and Compustat databases, respectively. Table 4.7 in Appendix 4.5.4 presents the complete list of variables used in the analysis that follows, their measurements, and relevant sources.

Table 4.8 in Appendix 4.5.4 provides some insights on the composition of the sample. Stock is the most common form of payment and is observed in around 58% of cases, followed by cash, which accounts for almost 42% of observations. Most of the announced bids are unchallenged, are classified as friendly and eventually end up being successfully completed. The sample includes a fair representation of deals that are intended for either business diversification (35%) or specialization (65%), classified on the basis of firms' two-digit SIC codes.[19] Only a few bids are rumored before they are announced or are for targets with a poison pill defensive provision in place or from bidders already owning a significant toehold in the target.

A few differences emerge comparing bid characteristics across cash and stock bids. First, the bid premium varies significantly across methods of payment and is higher, on average, for cash bids. This difference, approximately 10%, is statistically significant at the 1% level in a parametric t-test of the equality of means. In this respect, also Eckbo (2009) documents that among deal characteristics expected to affect the bid premium, the method of payment is one of the most important and, in particular, that premiums tend to be higher when cash is used. Then, stock bids are more common when the deal value is large, both in absolute and in relative terms (i.e. with respect to the bidder's market capitalization) as in Hansen (1987), or when the informational costs are lower and targets may thus be more willing to accept stock as a means of exchange, as for example whenever the transaction is friendly. On the contrary, cash bids are instead more common when targets would be less willing to accept stock as a means of exchange, as for example when informational problems are more severe, such as in tender offers, hostile bids and diversifying deals or when poison pills are in place. Moreover, cash bids are also more common when the informational costs for the bidder are lower, because it already has a toehold in the target, or when the transaction is made more complex by competition by a rival bidder. The expected synergies from the transaction of stock and cash bids are not statistically different, on average. Nor the frequency with which a bid is rumored varies significantly across methods of payment. Still, cash bids are more frequently withdrawn and are preceded, on average, by a higher run-up.

[19] A similar classification criterion on the basis of the first two digits of firms' SIC codes is adopted by Berger and Ofek (1995).

Table 4.2: Descriptive Statistics: bidder and target characteristics.

This table reports summary statistics for the bids included in the sample. Means and standard deviations are computed across the entire sample and conditional on the method of payment. In the last column, the result of a parametric t-test of the equivalence of means across methods of payment is presented. Table 4.7 describes the variables and their sources. The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	All Bids Mean St.Dev.	Cash Mean St.Dev.	Stock Mean St.Dev.	Cash-Stock Diff. (t-stat)
Target Market Cap. ($ mil.)	1,112	689	1,455	-766***
	4,259	1,387	5,571	(-5.02)
Target Market-to-Book	3.44	2.85	4.13	-1.28***
	10.05	10.59	9.34	(-2.81)
Bidder Market Cap. ($ mil.)	15,893	21,437	10,882	10,591***
	42,588	45,879	38,741	(5.76)
Bidder Market-to-Book	4.03	3.71	4.39	-0.68
	17.97	21.12	13.72	(-0.80)
Bidder Investment Opportunities	0.09	0.08	0.10	-0.02***
	0.11	0.08	0.13	(-3.41)
Bidder Leverage	0.22	0.23	0.20	0.03***
	0.19	0.20	0.19	(2.72)
Bidder Cash Holdings	0.10	0.10	0.11	-0.01**
	0.13	0.11	0.15	(-2.40)
Bidder Cash Flows	0.10	0.13	0.07	0.06***
	0.37	0.43	0.29	(3.79)

Table 4.2 summarizes then the firm characteristics of bidders and targets in our sample conditional on the method of payment they self-select into. As expected, bidder size is, on average, considerably larger than the size of targets and the average target involved in a stock bid is significantly larger than the average one that is offered cash. Stock bids usually involve bidders that need to preserve their cash, as they operate with higher cash holdings or have more investment opportunities. The same pattern is observable in the average market-to-book of target firms. Stock bids are targeted to firms with higher market- to-book ratios, consistent with bidders' greater concern of overpayment when market valuations are high relative to book values. Cash bids usually involve instead larger bidders, with better access to credit, as proxied by their higher leverage and their

greater cash flows generating power.

4.3.2 Methodology

We implement a multivariate analysis that controls for deal- and firm-specific attributes that, individually or in interaction with firm opacity, are expected to drive the determination of the observed method of payment and the bid premium. In particular, we consider the choice of the method of payment and the level of the bid premium as jointly determined as part of the optimal bidding strategy. As such, they are both endogenous variables, and then, we propose to model their determination in the form of a simultaneous equation system with a dummy variable identifying the method of payment, MP, and the bid premium, PRM, as the two endogenous variables:

$$\begin{aligned} &Pr(MP = stock|PRM, X_{MP}) = \Phi(\gamma_{MP}PRM + \beta_{MP}X_{MP} + \epsilon_{MP}) \\ &PRM = \gamma_{PRM}MP + \beta_{PRM}X_{PRM} + \epsilon_{PRM} \end{aligned} \tag{4.38}$$

Our approach resembles that used by Boone and Mulherin (2007) in their analysis of wealth effects in auctions versus negotiations and is grounded on the simultaneous equation models with qualitative and continuous dependent variables described by Maddala (1983).[20] More specifically, in the first equation the dependent variable is a dummy variable equal to 1 for stock bids, and 0 otherwise. In the second equation, instead, the dependent variable is the percentage amount by which the offering price exceeds target's undisturbed stock price (i.e. four weeks before the announcement).[21] X_{MP} and X_{PRM} are vectors including the opacity of the bidder, the opacity of the target and the set of control variables respectively related to the choice of the method of payment and the bid premium. Individual effects in isolation would, in fact, fail to take into account that many factors concur to jointly determine the observed method of payment. For example, just considering firm opacity, not only does the method of payment eventually observed depend on the opacity of one counterparty, but rather on the interplay of the - sometimes conflicting - effects of bidder and target opacities. Whether a cash or stock bid is observed, in fact, depends on which effect prevails. Moreover, there are many other factors

[20] Other examples of previous applications of simultaneous equation models in similar settings with qualitative dependent variables include Hansen (1986) studying the revenue equivalence of sealed bids versus open auctions, Smith (1987) providing a comparative analysis of the proceeds in competitive versus negotiated securities offerings, Demsetz and Lehn (1985) focusing on the link between a firm performance and its ownership status, Comment and Schwert (1995) and Officer (2003) assessing the wealth effects of respectively poison pills and termination provisions

[21] See Eckbo (2009) for a review of the different proxies for the bid premium used in the literature

that together with firms opacity could play a role in the determination of the method of payment, as for example the fact that some bidders might not have enough liquid resources to make a cash bid or may be credit constrained. The set of controls for bid characteristics includes: deal materiality, the target stock price run-up, a dummy variable to identify tender offers, a dummy variable to identify hostile bids, a dummy variable to identify bids involving firms operating in different industries and a dummy variable to identify bids occurring during a merger wave, defined in Harford (2005).[22] At the firm level, Fama-French 12-industries dummy variables are used to proxy for common unobserved bidder and target characteristics. Macro variables include Shiller's P/E index and the C&I loan spread that Harford (2005) indicates as two main drivers of M&A activity.

4.3.3 Firm opacity

Unfortunately, opacity, our cross-sectional conditioning variable, is not directly observable. Still, a firm's equity trading properties - and its liquidity in particular - can reflect the nature of the information available to market participants on the value of the firm. Based on this premise, we assume that the information asymmetry faced by counterparties in a deal is to some extent correlated with that of other outsiders and we rely on the adverse selection component extracted from existing measures of liquidity to proxy for firm opacity.[23]

However, the concept of liquidity is tightly and elusively interconnected to asymmetric information. Indeed, according to Hasbrouck (2009), while liquidity is intended as the ability to trade promptly and with little or no price impact and it is then closely related to the extent of uncertainty over the value of the asset, there is no single measure that captures all of its dimensions. As a consequence, a possible concern is that every single potential proxy of liquidity is driven by adverse selection, but not exclusively so. We then design an index of firm opacity by capturing on the first principal component the common cross-sectional variation of six different constituents: (i) the illiquidity measure of Amihud (2002), (ii) the volume $-$ return autocorrelation of Llorente et al. (2002), (iii) the probability of informed trading of Easley et al. (1996), (iv) the adverse selection

[22]Specifically, we consider a deal occurring during a merger wave if in the same period we assess an exceptional concentration of merger activity within the industry of either the target, the bidder, or both. Details on the construction of the latter variable are available upon request.

[23]Adverse selection risk is the risk of facing better-informed counterparties when trading a specific stock. It increases with firm opacity. The link between equity trading characteristics and information is indirectly validated by Chae (2005), who documents that measures of market microstructure are significantly affected by announcements of corporate events, including M&As.

component of the proportional effective spread of Roll (1984), (v) the reversal coefficient of Pastor and Stambaugh (2003), and (vi) the Amivest liquidity ratio of Cooper et al. (1985) and Amihud et al. (1997). The intuition is that combining broader liquidity measures with more informational proxies on their first principal component minimizes the likelihood that these measures are connected to non-informational liquidity. Our approach replicates that of Bharath et al. (2009), who form an index to study the impact of a firm's private information on capital structure decisions.

Our index is computed for each bidder and each target in the year preceding the bid announcement. Relevant loadings on the individual components of the index are extracted by principal component analysis of our index constituents in each year for all firms with data available from CRSP.[24]

Specifically, we estimate the first principal component of the correlation matrix of the standardized index constituents and then, for each firm, we form the index of firm opacity by combining our standardized proxies for firm opacity with the corresponding contemporaneous loadings.[25] Higher values of the index are associated with higher opacity for the specific firm in the given year. A detailed description of the constituents of our index, how it is constructed, and its main properties is provided in the Appendix, where we also present some robustness tests to validate its use in our empirical analysis. According to our index, the opacity of firm i in year y is computed on the basis of our six index constituents x standardized across all firms in the given year, as

$$Index_{i,y} = \sum_{j=1}^{6} w_{j,y}\bar{x}_{i,y} \qquad \text{where} \qquad w_{j,y} = PC(\bar{x}_{i,y}) \tag{4.39}$$

Higher values of the index are associated with higher opacity for the specific firm in the given year. In each year, as well as overall, the mean index value across all firms on CRSP is zero by construction. Still, as reported in Table 4.3 bidders and targets in our sample are on average more transparent than the average firm in CRSP. Moreover, targets tend to be on average relatively more opaque than bidders, consistent with their relative size.

[24] All firms with data available on CRSP are considered in the analysis, since the cross section of firms in our sample of bidders and targets over single years is limited and not homogeneous. The broader scope improves the efficiency of the principal component analysis. On average, 40% of cross-sectional variance is accounted for by the first principal component and in most years only the first eigenvalue is larger than one. Moreover, the elements of the first eigenvector are mostly positive, confirming that each constituent adds positively to the index.

[25] Index constituents are standardized across all firms with data available on CRSP in a given year, since a broader scope improves the efficiency of the principal component analysis. The results of the analysis under alternative standardizations at the industry level or across firm size quartiles are available upon request.

Table 4.3: Descriptive Statistics: bidder and target opacity.

This table reports summary statistics for the index of opacity for the targets and bidders included in our sample. The index is computed as in Bharath et al. (2009). Table 4.7 describes the variables and their sources.

	Obs.	Mean	St. Dev.	$p-25^{th}$	Median	$p-75^{th}$
Targets	2,242	-0.48	0.83	-0.86	-0.49	-0.12
Bidders	2,046	-1.41	1.02	-2.09	-1.27	-0.71
All	4,189	-0.86	1.01	-1.34	-0.72	-0.29

4.3.4 Results

We assess the impact of firm opacity on the simultaneous determination of the bid premium and the method of payment by consistent estimation of our system of equations with a two-stage probit least squares regression analysis. In particular, in the first stage we regress the two dependent variables, the method of payment and the bid premium on their corresponding sets of exogenous variables. Then, in the second stage regressions, we use the fitted values for a given dependent variable as an explanatory variable for the other one. Since the choice of the method of payment is modeled as a probit model, standard errors in the second stage are estimated according to Maddala (1983).[26] Identification of the system requires that X_{MP} and X_{PRM} do not coincide. To assure identification, then, the set of regressors for the method of payment X_{MP} will not include the pre-announcement run-up of the stock price of the target that we use to instrument for the bid premium. We consider, in fact, the target's stock price run-up a potential driver of the bid premium while we deem it unrelated to the method of payment. Our intuition is supported by Eckbo (2009) and Betton et al. (2014) that show how a high pre-announcement run-up is positively associated with a higher bid premium and by the fact that in our sample the relative frequency with which stock and cash bids occur is not affected by the pre-announcement run-up in target stock prices. Univariate tests of equality of means confirm the null hypothesis that the proportion of stock bids does not vary for different levels of run-up, not across subsamples of deals where run-up is above or below its cross-sectional median nor where it is positive or negative. Analogously, the dummy variable that identifies tender offers is excluded from the set of regressors of the bid premium X_{PRM} and used to instrument for the method of payment. We consider,

[26] More specifically, to the compute the proper estimates we follow Keshk (2003)

in fact, the decision to realize the transaction by means of a tender offer to be closely tied to the choice of the method of payment, in line with Huang and Walkling (1987) and Berkovitch and Khanna (1991) who show that the likelihood of stock bids is lower in tender offers and our summary statistics in Table 4.8, and mostly unrelated to the level of the bid premium after controlling for the differences across methods of payment.

Table 4.4 summarizes the coefficients and the corresponding t-statistics (in parenthesis) for the second stage regressions under three alternative model specifications.[27] We start with a baseline specification (I) including only the controls at the bid and macro level; then we include an interaction term in the method of payment equation of our second specification (II) to assess the effect of target opacity for different levels of deal materiality, following Hansen (1987); and finally, in our third specification (III) we complete the set of exogenous variables by adding firm level controls.

Our analysis documents that the opacity faced by the bidder in assessing the value of the target is a significant driver of the choice of the method of payment. Indeed, consistent with the predictions of our model, the likelihood of a stock bid increases with the opacity of the target. The coefficient of target opacity in the equation for the method of payment in our baseline specification (I) is positive, but the magnitude of the effect varies with deal materiality. Indeed, in our model specifications (II) and (III), the interaction term between target opacity and deal materiality, which captures the increasing concern of overpayment as bid size grows, is always positive and significant at the 5% level. This suggests that it is when the transaction is sufficiently material that stock bids are preferred to alleviate the overpayment concerns associated with opaque targets.[28] The economic magnitude of this effect is substantial. According to the marginal effects based on our model specification (III), a one standard deviation increase in target opacity would increase, for example, the probability of choosing a stock payment by roughly 2.4% for a deal where the size of the target accounts for one-third of the combined merged entity, i.e. when the bidder is twice as large as the target, or by as much as 5.6% vice-versa.[29] This result is in line with the predictions of our model, which are based on the same risk-sharing rationale of Hansen

[27] The individual effects of control variables are omitted from the table for the sake of space and are available upon request

[28] Analyzing a smaller subsample of deals which does not include observations belonging to the two smallest deciles with respect to deal materiality, i.e. deals for which the target accounts for less than 2.5% of the combined merged entity, leads to equivalent conclusions. The coefficient of target opacity in the equation for the method of payment in our baseline specification (I) is positive and significant at the 1% level. Results are available upon request.

[29] In particular, the marginal effect of target opacity on the probability of a stock bid is 0.0289, significant at the 10% level, when deal materiality is fixed at 0.33 and equals 0.0313, significant at the 5% level for deal materiality equal to 0.66

Table 4.4: Model Estimation.

This table reports the coefficients and t-statistics (in parentheses) for the simultaneous estimation by means of Keshk (2003)'s two-stage probit least squres of our system of equations (4.38) modeling the method of payment (first column) and the bid premium (second column). Only the estimates of the second stage regressions are reported, first stage regressions are available upon request. In the first equation, the dependent variable, MP, is a dummy variable equal to 1 for stock bids, and 0 otherwise. In the second equation, the dependent variable is the percentage amount by which the offering price exceeds target's stock price four weeks before the announcement. The set of controls for bid characteristics includes: deal materiality, the target's stock price run-up, a dummy variable to identify tender offers, a dummy variable to identify hostile bids and a dummy variable to identify bids involving firms operating in different industries and a dummy variable to identify bids occurring during a merger wave, defined as in Harford (2005). At the firm level, Fama-French 12-industry dummy variables are used to control for bidder characteristics. Macro variables include Shiller's P/E index, and the C&I Loan spread. The target stock price run-up is used to instrument for the bid premium and is thus excluded from the controls in the first regression, while the dummy variable to identify tender offers is used to instrument for the method of payment and thus is excluded from the second regressio. A description of the variables used in the analysis and the relevant sources is provided in Table 4.7. The individual effects of control variables are omitted from the table for the sake of space and are available upon request. The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	(I)		(II)		(III)	
	MP	PRM	MP	PRM	MP	PRM
Target Opacity	0.0209	0.0462***	0.0631	0.0462***	-0.1490*	0.0483***
	(0.33)	(3.29)	(0.82)	(3.29)	(-1.82)	(3.40)
x Materiality	-	-	0.6099**	-	0.7798**	-
	-	-	(2.04)	-	(2.39)	-
Bidder Opacity	-0.0552	-0.0212*	-0.0792	-0.0212*	-0.0913	-0.0213*
	(-0.99)	(-1.73)	(-1.40)	(-1.73)	(-1.54)	(-1.72)
Controls: Bid	Yes	Yes	Yes	Yes	Yes	Yes
Controls: Firm	No	No	No	No	Yes	Yes
Controls: Macro	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	1646	1646	1646	1646	1646	1646

(1987) and in contrast with the use of cash bids as a signaling device to deter potential competitors' bids for more opaque targets, as instead documented by Chemmanur et al. (2009).

We only find weak evidence that the likelihood of a stock bid is overall negatively affected by the opacity of the bidder. Indeed, while this effect is of the expected sign, it does not reach statistical significance at the 10% level in any of our specifications. As mentioned before, the sharp prediction of a negative correlation depends in our model on assumptions about the distribution of prior beliefs. Because different channels display different effects of the components of bidder opacity on the likelihood of a stock offer, the small significance of this correlation as displayed by the data is not surprising and does not invalidate the model's intuition.

We do not find evidence on the use of cash bids as a signaling device to deter potential competitors' bids for more opaque targets, as do Chemmanur et al. (2009), but we find instead, as predicted by our model, that higher bid premiums are associated with target opacity. In this respect, our findings can also be interpreted in light of the preemptive bidding rationale of using cash bids as a signaling device to deter potential competitors' bids for more opaque targets, as suggested by Fishman (1989) and consistent with the evidences provided by Laamanen (2007) and Chatterjee et al. (2012). In particular, in all our model specifications, target opacity is positively related with the level of the bid premium. The corresponding coefficients are always positive and statistically significant at the 1% level. Also this effect is economically relevant as, for example, according to the estimates in our specification (III), for a one standard deviation increase in target opacity the bid premium would increase, on average, by 4.0% in absolute terms. Or, by approximately $ 44.5 million considering the average target in our sample.[30]

Our analysis then documents that, after controlling for the method of payment, the opacity of the bidder is related to lower premiums. While this is not direct evidence in favor of our hypothesis that bidder opacity is negatively correlated with premiums conditional on stock offers (something which cannot be tested given our specification), it is consistent with our theoretical model where bidder opacity makes it hard for high-type targets to sustain stock-pooling equilibria, which carry the largest premiums while bidder opacity is associated with higher premiums in separating equilibria through cash offers only. According to the alternative argument of Myers and Majluf (1984) and Rhodes-Kropf and Viswanathan (2004), in fact, when the bidder is opaque, targets are more likely to overestimate synergies and the ensuing increased probability of bid success allows then more opaque bidders to signal their value by offering less at relatively lower cost. In particular, we find, in all our model specifications, that bidder opacity is negatively correlated with the level of the bid premium. The corresponding coefficients are always

[30] Considering instead the median target size the corresponding increase in the bid premium would be equal to $ 9 million.

negative and statistically significant at the 10% level. Also this effect is economically relevant as, for example, according to the estimates in our specification (III), for a one standard deviation increase in bidder opacity the bid premium would decrease, on average, by 2.2% in absolute terms. Or, by approximately $ 24.5 million considering the average deal in our sample.[31]

Through this conditional effect, bidder opacity then indirectly contributes to the choice of stock as the method of payment since it lowers a bidder's expectation of a stock bid premium, affecting the gap between the premiums that the bidder anticipates to offer under alternative payment regimes. As a consequence, whether a cash or stock bid is observed depends on which effect prevails. Indeed, a relatively lower anticipated premium under stock payment when a bidder is more opaque would make this type of bid more likely, offsetting the otherwise negative effect of bidder opacity on the use of stock. This indirect contribution of bidder opacity to the choice of the method of payment may then provide an explanation why we only find weak evidence that the likelihood of a stock bid is overall negatively affected by the opacity of the bidder, as predicted by our model.

Evidence of the effect of control variables is generally as expected and in line with Boone et al. (2014). Preference for a stock transaction increases with deal materiality. Consistent with the evidence of Faccio and Masulis (2005), the corresponding coefficient is positive and strongly statistically significant. Analogously, in line with, among others, Schwert (2002) and Faccio and Masulis (2005), a preference for stock bids is found to be significantly higher for deals that involve firms in the same industry. These classes of deals are, in fact, more likely to include transactions in which asymmetric information concerns are lesser and in which the target's shareholders are more likely interested in maintaining a stake in the merged entity. As a consequence, these deals are also associated with significantly lower bid premiums. Then, consistent with McNamara et al. (2008), Chidambaran et al. (2010), and Chatterjee et al. (2012), our analysis suggests a significant preference for stock payments in deals that are part of a merger wave or that occur in periods of strong investment sentiment or low capital liquidity, as proxied respectively by Shiller's P/E index and the C&I Loan spread. The corresponding coefficients are consistently significant. Finally, as expected, run-up is positively and significantly associated with the bid premium, consistent with Betton et al. (2014) and the likelihood of a stock bid is lower in tender offers, consistent with Huang and Walkling (1987) and Berkovitch and Khanna (1991).

[31] Considering instead the median target size the corresponding increase in the bid premium would be equal to $ 5 million.

4.4 Conclusion

Our theoretical model and empirical tests show the joint effect of target and bidder opacity on the simultaneous determination of the method of payment and the bid premium in M&A transactions. The theoretical part provides a few testable hypotheses on the relationship between opacity and the transaction's observable characteristics. The empirical analysis takes these predictions to the data, using a sample of M&A bids by and for U.S. publicly listed firms over the period 1985-2011.

Overall, our results lean on the observation that the concern of overpayment leads bidders to select stock bids to benefit from contingent pricing and risk sharing. Modeling and testing jointly and directly the impact of both target and bidder opacity on bid characteristics, we are able to observe that the preference for stock bids increases with the opacity of the target for bids of substantial materiality, consistent with the adverse selection rationale of Hansen (1987). Moreover, we observe that premiums are higher for cash bids and increase with the opacity of the target, but they are also negatively related to bidder opacity. The first results are consistent with arguments by Fishman (1989) and Chatterjee et al. (2012), while the latter may support the arguments of Myers and Majluf (1984) and Rhodes-Kropf and Viswanathan (2004).

Our direct investigation of the implications of firm opacity on the realization of an M&A deal sheds light on the motives underlying the observed bidding behavior and the efficiency of the market for corporate control, by quantifying the impact of the entailed informational frictions on managerial decisions and negotiations. Moreover, it may provide foundations for the prominent role played by financial intermediaries acting as advisors and the continuous effort to design market devices to convey relevant information. In this respect, our results are related to those of Kesner et al. (1994), who take an agency theory perspective and evaluate how the interests of – respectively – bidders, targets, and their advisors reflect on bidding behavior. Furthermore, our results complement those of Reuer et al. (2012), who document how a target's association with a prominent investment bank, venture capitalist, or alliance partner conveys valuable information and positively affects the bid premium.

A natural extension of our analysis would then be to adapt the proposed model for the choice of the method of payment and the bid premium to study the information content of a deal's announcement and show the implications of how much is paid and how for deal success and shareholder value creation. In particular, it would be interesting to look further into the different types and quality of information on the bidder and the target that stock and cash bids respectively convey and, on these premises, shed new light on the

role played by the choice of the most appropriate method of payment as a determinant of value creation for shareholders.

4.5 Appendices

4.5.1 Equilibrium Refinements

Some PBEs may be sustained only by the target holding an "unreasonable" posterior belief after receiving an out-of-equilibrium message. In particular, we consider it unreasonable, for the target, to assign, upon observing m, positive probability to a type of bidder which is better off with the outcome of its equilibrium action than with any outcome it might receive by choosing to play m, given its assessment of the target's rationality.

In order to translate this into formal language,[32] let us consider a candidate equilibrium *path* $(\mathbf{m}^*, \hat{\mathbf{r}}^*)$. Here $\mathbf{m}^* \in M^{\Theta_B}$ is the bidder's equilibrium strategy and $\hat{\mathbf{r}}^* \in \{Y, N\}^{\Theta_T \times \mathbf{m}^*(\Theta_B)}$ is the restriction to the domain of equilibrium messages of the target's equilibrium strategy. Let us also define, given m^* and the prior belief p^T, the set of Bayes*-consistent posterior beliefs (Ba*) to be the set of beliefs $\boldsymbol{\mu} \in \Delta(\Theta_B)^M$ such that $\boldsymbol{\mu}(m)$ is derived from p^T via Bayes' rule whenever $m \in \mathbf{m}^*(\Theta_B)$.

We say that the target is *-rational (and denote such event by R_T^*) if, for all $\theta_T \in \Theta_T$ and for all $m \in M$, it chooses an action which is a best reply to m given *some* belief μ consistent with $\mu = \boldsymbol{\mu}(m)$, $\boldsymbol{\mu} \in \text{Ba}^*$. The set of strategies $\mathbf{r} \in \{Y, N\}^{\Theta_T \times M}$ which are compatible with R_T^* is denoted by $r(R_T^*)$. Similarly, we say that the bidder is *-rational (and denote such event by R_B^*) if, for all $\theta_B \in \Theta_B$, the bidder chooses $m(\theta_B)$ to be a best reply to *some* conjecture about the target's behavior, given that the target will choose $r \in \hat{\mathbf{r}}^*(\Theta_T, m)$ whenever $m \in \mathbf{m}^*(\Theta_B)$.

By definition[33] a candidate equilibrium path cannot be inconsistent with either R_T^* or R_B^*. For an event E and each $i \in \{B, T\}$, let $\text{B}_i(E)$ denote the event "player i initially believes E". Consider the event $R_B^* \cap \text{B}_B(R_T^*)$, which reads "the bidder is *-rational and (initially) believes that the target is *-rational". Such event constrains the bidder to choose $m(\theta_B)$ taking into account that the target will respond to the message with an action consistent with some strategy in $r(R_T^*)$. Thus, a message m by a type-θ_B bidder is consistent with

[32] This appendix is based on the epistemic analysis of forward-induction solution concepts contained in Battigalli and Siniscalchi (2002); see also Battigalli and Siniscalchi (2003) and Battigalli (2006).

[33] In this paper, the reference equilibrium concept is the Perfect Bayesian Equilibrium defined in Definition 1. What follows, though, holds even when the reference equilibrium concept is self-confirming equilibrium; Battigalli and Siniscalchi (2002).

$R_B^* \cap \mathrm{B}_B(R_T^*)$ if and only if there exists a $\tilde{\mathbf{r}} \in r(R_T^*)$ such that

$$m \in \arg\max_{\tilde{m} \in M} \mathbb{E}_{p^B}\left[u_B(\tilde{m}, \tilde{\mathbf{r}}(\theta_T, \tilde{m}), \theta_B, \theta_T\right] \tag{4.40}$$

If m is an out-of-equilibrium message (i.e. $m \notin \mathbf{m}^*(\Theta_B)$) and there is no such strategy $\tilde{\mathbf{r}}$, it means that the expected payoff obtained by the type-θ_B bidder by playing its equilibrium message $\mathbf{m}^*(\theta_B)$ is strictly larger than what it can obtain by playing m. In this case, we say that m is *equilibrium-dominated*. For each message m, we let $\Theta_B^*(m)$ be the set of types of the bidder for which m is *not* equilibrium-dominated.

For an event E an player $i \in \{B, T\}$, the event $\mathrm{SB}_i(E)$ indicates that player i *strongly believes* E: it initially believes E and keeps doing so throughout the game, as long as what it observes is compatible with E. Consider now the event $R_T^* \cap \mathrm{SB}_T\left(R_B^* \cap \mathrm{B}_B(R_T^*)\right)$: the target is *-rational and strongly believes the bidder to be *-rational as well as to believe in the target's own *-rationality.

On its own, the event $\mathrm{SB}_T\left(R_B^* \cap \mathrm{B}_B(R_T^*)\right)$ implies that the target does not attribute to the type-θ_B bidder the choice of an equilibrium-dominated message. Because Bayes-consistency leaves the target free in forming posterior beliefs after non-equilibrium messages, it might be that a target assigns positive probability to a type-message pair (θ_B, m) such that m is equilibrium-dominated for θ_B.

Definition 8. *A PBE* $(\mathbf{m}^*, (\boldsymbol{\mu}^*, \mathbf{r}^*))$ *fails the* Equilibrium Dominance Test (EDT) *if there is* $m \in M$ *such that* $\boldsymbol{\mu}^*(\Theta_B^*(m) \mid m) < 1$.

In our paper, it is often the case that many PBEs exist giving rise to the same equilibrium path and differing only in the posterior beliefs. Very often, removal of PBEs that do not pass the EDT ends up in curbing the amount of equilibria who share a given outcome. In other words, for a given equilibrium path, one can find at least one assessment which sustains such path as the outcome of a PBE which passes the EDT. In general, though, the event $\mathrm{SB}_T\left(R_B^* \cap \mathrm{B}_B(R_T^*)\right)$ might make some equilibrium outcomes unsustainable. To see this, define, for all $A \subseteq \Theta_B$,

$$\mathrm{Ba}^*(A) := \{\boldsymbol{\mu} \in \mathrm{Ba}^* \,:\, \forall m \in M\,,\, \boldsymbol{\mu}(A \mid m) = 1\} \tag{4.41}$$

Compatibility with the event $R_T^* \cap \mathrm{SB}_T\left(R_B^* \cap \mathrm{B}_B(R_T^*)\right)$ requires that the target's response to message m is a best reply to some posterior beliefs that assigns probability 1 to m having been sent by a type of bidder for which m is not equilibrium dominated. That is, it is required that $\mathbf{r}(\theta, m)$ be a best response to some belief μ such that $\mu = \boldsymbol{\mu}(m)$ and $\boldsymbol{\mu} \in \mathrm{Ba}^*(\Theta_B^*(m))$. As before, then, denote by $r(R_T^* \cap \mathrm{SB}_T\left(R_B^* \cap \mathrm{B}_B(R_T^*)\right))$ the strategies satisfying such requirements.

Finally, consider the event $R_B^* \cap \mathrm{B}_B\left(R_T^* \cap \mathrm{SB}_T\left(R_B^* \cap \mathrm{B}_B(R_T^*)\right)\right)$. Fix a type $\theta_B \in \Theta_B$. A message m by θ_B is *not* compatible with this event if there is another message m' that gives the type-θ_B bidder a larger payoff irrespectively of the target's response to m', provided such response is consistent with a strategy $\mathbf{r} \in r(R_T^* \cap \mathrm{SB}_T\left(R_B^* \cap \mathrm{B}_B(R_T^*)\right))$. It is possible that some PBE prescribes an equilibrium path in which, for some type θ_B, the type-θ_B bidder is prescribed an equilibrium message that makes the equilibrium path not compatible with $R_B^* \cap \mathrm{B}_B\left(R_T^* \cap \mathrm{SB}_T\left(R_B^* \cap \mathrm{B}_B(R_T^*)\right)\right)$. The following definition parallels the original one of Cho and Kreps (1987).

Definition 9. *A PBE* $(\mathbf{m}^*, (\boldsymbol{\mu}^*, \mathbf{r}^*))$ *fails the* Intuitive Criterion Test (ICT) *if there exist* $m \in M$ *and* $\theta_B \in \Theta_B$ *such that*

$$U_{\theta_B}^*(\mathbf{m}^*(\theta_B)) < \min_{\mathbf{r} \in r(R_T^* \cap \mathrm{SB}_T(R_B^* \cap \mathrm{B}_B(R_T^*)))} \mathbb{E}_{p^B}\left[u_B(m, \mathbf{r}(\theta_T, m), \theta_B, \theta_T\right] \tag{4.42}$$

Thus, an equilibrium fails the ICT test if the message it prescribes to some type of bidder turns out to be worse than some out-of-equilibrium message, once the bidder considers that the target will reply to the latter with a response consistent with the target assigning zero probability to the out-of-equilibrium message having been sent by a type of bidder for which such message would be equilibrium-dominated.

4.5.2 Equilibrium Analysis

By the following proposition we identify the minimal stock offers that, given a certain updating rule $\mu \in \Delta(\Theta_B)^M$, are accepted, respectively, by the low-type and the high-type target.

Proposition 10. *For all* $\theta_T \in \Theta_T$ *and* $\mu \in \Delta(\Theta_B)^M$*, the map* $\Phi^{\theta_T} : [0,1] \to \mathbb{R}_+$ *defined in (4.21) has a unique fixed point* $f^{\theta_T}(\mu) \in (0,1)$*. Moreover,* $f^{\overline{\theta}_T}(\mu) \geq f^{\underline{\theta}_T}(\mu)$*.*

Proof. By assumption (A3), $\mu\left(\overline{\theta}_B \mid (S, \cdot)\right)$ is continuous and weakly increasing. Then, for all $\theta_T \in \Theta_T$,

$$\Phi^{\theta_T}(\cdot) := \frac{\theta_T}{\mathbb{E}_{\mu(S,\cdot)}\left[\theta_B\right] + w(\theta_T)} = \frac{\theta_T}{\mu\left(\overline{\theta}_B \mid (S, \cdot)\right) V_B + \underline{\theta}_B + w(\theta_T)}$$

is continuous and weakly decreasing. Moreover, for all $f \in [0,1]$,

$$\Phi^{\theta_T}(f) = \left[\frac{\theta_T}{\overline{\theta}_B + w(\theta_T)}, \frac{\theta_T}{\underline{\theta}_B + w(\theta_T)}\right] \subset (0,1)\,.$$

Therefore there exists, for all $\theta_T \in \Theta_T$ and $\mu \in \Delta(\Theta_B)^M$, a unique value $f^{\theta_T}(\mu)$ that solves $f = \Phi^{\theta_T}(f)$. Assumption (A2) implies that, for all $\nu \in \Delta(\Theta_B)$,

$$\frac{\overline{\theta}_T}{\underline{\theta}_T} \geq \frac{\mathbb{E}_\nu[\theta_B] + w(\overline{\theta}_T)}{\mathbb{E}_\nu[\theta_B] + w(\underline{\theta}_T)}, \tag{4.43}$$

from which it follows that, for all $\mu \in \Delta(\Theta_B)^M$ and all $f \in [0,1]$, $\Phi^{\overline{\theta}_T}(f) \geq \Phi^{\underline{\theta}_T}(f)$. Then it must be that $f^{\overline{\theta}_T}(\mu) \geq f^{\underline{\theta}_T}(\mu)$. □

We first prove Lemma (3).

Proof of Lemma (3). For all $\theta_T \in \Theta_T$, $\frac{\theta_T}{\overline{\theta}_B + w(\theta_T)} < \frac{\theta_T}{\underline{\theta}_B + w(\theta_T)}$, implying that

$$\forall \theta_T \in \Theta_T \quad , \quad f^{\theta_T}(\underline{\theta}_B) > f^{\theta_T}(\overline{\theta}_B) \tag{4.44}$$

Then, by (A2),

$$\frac{\overline{\theta}_T}{\underline{\theta}_T} \geq \frac{\overline{\theta}_B + w(\overline{\theta}_T)}{\underline{\theta}_B + w(\underline{\theta}_T)} \implies f^{\overline{\theta}_T}(\overline{\theta}_B) \geq f^{\underline{\theta}_T}(\underline{\theta}_B) \tag{4.45}$$

Using (4.44) and (4.45) we obtain (4.28). □

Now we prove Lemma (2).

Proof of Lemma (2). (a) For all $\theta_B \in \Theta_B$, the participation share constituting the low stock offer is found, by (4.21), to be

$$f^{\underline{\theta}_T}(\theta_B) = \frac{\underline{\theta}_T}{\theta_B + w(\underline{\theta}_T)} \tag{4.46}$$

Using (4.24), the expected value of such offer to the bidder is

$$\begin{aligned} & p^B\theta_B + (1-p^B)\left[\left(\frac{\theta_B + \Delta(\underline{\theta}_T)}{\theta_B + w(\underline{\theta}_T)}\right)(\theta_B + w(\underline{\theta}_T))\right] - \theta_B \\ = \; & p^B\theta_B + (1-p^B)\left(\theta_B + \Delta(\underline{\theta}_T)\right) - \theta_B \\ = \; & (1-p^B)\Delta(\underline{\theta}_T) \end{aligned}$$

which is the expected value of the low cash offer, as in (4.15).

(b) Similarly, the participation share of the high stock offer is

$$f^{\underline{\theta}_T}(\theta_B) = \frac{\overline{\theta}_T}{\theta_B + w(\overline{\theta}_T)} \tag{4.47}$$

and, by (4.23), the expected value of this offer is

$$\begin{aligned}
&\left(1-\frac{\overline{\theta}_T}{\theta_B+w(\overline{\theta}_T)}\right)(\theta_B+\mathbb{E}_{p^B}[w(\theta_T)])-\theta_B \\
=\;& \mathbb{E}_{p^B}[w(\theta_T)]-\overline{\theta}_T\left(\frac{\theta_B+\mathbb{E}_{p^B}[w(\theta_T)]}{\theta_B+w(\overline{\theta}_T)}\right) \\
>\;& \mathbb{E}_{p^B}[w(\theta_T)]-\overline{\theta}_T
\end{aligned}$$

the last line being the value of the high cash offer, as in (4.16). Notice that the values of the two high offers tend to coincide as $p^B \to 1$.

(c) Notice from formulas (4.15), (4.16), (4.23) and (4.24), that the expected payoffs from all offers are linear functions of p^B. By part (a) of this Lemma, the payoff from the low stock offer tends, as $p^B \to 0$, to $\Delta(\underline{\theta}_T)$. The payoff from the high stock offer tends, as $p^B \to 0$, to $w(\overline{\theta}_T)-\overline{\theta}_T\,\frac{\theta_B+w(\underline{\theta}_T)}{\theta_B+w(\overline{\theta}_T)}$ which, by assumption (A2), is strictly smaller than $\Delta(\underline{\theta}_T)$. As stated in part (b) of this Lemma, the high stock offer gives the bidder a larger payoff than the high cash offer which, for all $p^B \geq \pi_C$, gives a larger payoff than the low cash offer. This together with the linearity of such payoffs proves part (c) of the Lemma.

(d) Part (c) of this Lemma immediately yields that, for all $\theta_B \in \Theta_B$, $\pi_S(\theta_B) < \pi_C$. Moreover, from (4.23) one can see that the payoff from the high stock offer grows in p^B at rate $(1-f^{\overline{\theta}_T}(\theta_B))W$. As $f^{\overline{\theta}_T}(\underline{\theta}_B) > f^{\overline{\theta}_T}(\overline{\theta}_B)$, the slope of the payoff line is smaller for $\underline{\theta}_B$ than for $\overline{\theta}_B$, implying that $\pi_S(\underline{\theta}_B) < \pi_S(\overline{\theta}_B)$. □

The next Lemma excludes the possibility of an equilibrium where the two types of bidder propose distinct stock offers.

Lemma 11. *There is no PBE such that, for all $\theta_B \in \Theta_B$, $\mathbf{m}^*(\theta_B) \in S \times [0,1]$ and $\mathbf{m}^*(\overline{\theta}_B) \neq \mathbf{m}^*(\underline{\theta}_B)$.*

Proof. Assume, towards a contradiction, that such a PBE existed. By definition, for all $\theta_B \in \Theta_B$, $\boldsymbol{\mu}^*(\theta_B \mid \mathbf{m}^*(\theta_B)) = 1$ (full recognition of the bidder). As $f \mapsto \boldsymbol{\mu}^*\left(\overline{\theta}_B \mid (S,f)\right)$ is required to be increasing by (A3),

$$\boldsymbol{\mu}^*\left(\overline{\theta}_B \mid \mathbf{m}^*(\overline{\theta}_B)\right) = 1 > 0 = \boldsymbol{\mu}^*\left(\overline{\theta}_B \mid \mathbf{m}^*(\underline{\theta}_B)\right)$$

implies $\mathbf{m}^*(\overline{\theta}_B) > \mathbf{m}^*(\underline{\theta}_B)$. By Lemma (3), this can only happen if $\mathbf{m}^*(\overline{\theta}_B) = f^{\overline{\theta}_T}(\overline{\theta}_B)$ and $\mathbf{m}^*(\underline{\theta}_B) = f^{\underline{\theta}_T}(\underline{\theta}_B)$. This in turn requires $p^B \geq \pi_S(\overline{\theta}_B)$ and $p^B \leq \pi_S(\underline{\theta}_B)$, which is impossible by part (d) of Lemma (2). □

The next lemma excludes the possibility of a PBE where the high-type bidder offers stock and the low-type bidder offers cash.

Lemma 12. *There is no PBE such that* $\mathbf{m}^*(\underline{\theta}_B) \in C \times \mathbb{R}_+$ *and* $\mathbf{m}^*(\overline{\theta}_B) \in S \times [0,1]$.

Proof. Once again, suppose to have such a PBE. For all $\theta_B \in \Theta_B$ the payoff from a stock offer is decreasing in the participation share f which, in turn, is decreasing in the target's belief that the bidder's type is $\overline{\theta}_B$ at (S, f). For every $p^B \in (0,1)$, the best stock offer of a type-$\underline{\theta}_B$ bidder features a participation share that exceeds the one offered by the type-$\overline{\theta}_B$ bidder in equilibrium. Thus, if it deviated to such offer, the type-$\underline{\theta}_B$ bidder would be mistakenly considered a type-$\overline{\theta}_B$ bidder. Its payoff would then be larger than the one it could obtain from the same offer under full recognition, which, in turn, is weakly larger than the payoff from the best cash offer. It follows that such a deviation is always profitable for the type-$\underline{\theta}_B$ bidder and, therefore, that there cannot be any such PBE. □

We can finally prove the main result on separating equilibria of our model.

Proof of Proposition (4). Fix $p^B \in (0,1)$ and let, for simplicity, c^* be the best cash offer at p^B. Let also s^* be the best stock offer for a type-$\underline{\theta}_B$ bidder at p^B. We look for an updating rule $\boldsymbol{\mu}^*$ such that $\boldsymbol{\mu}^*$ satisfies (A3) and $(\mathbf{m}^*, (\boldsymbol{\mu}^*, \mathbf{r}^*))$ is a PBE, where $\mathbf{m}^*(\overline{\theta}_B) = c^*$, $\mathbf{m}^*(\underline{\theta}_B) = s^*$ and $\mathbf{r}^*$ is the appropriate response.

A type-$\underline{\theta}_B$ bidder has no incentive to deviate from s^* to c^* as s^*, guaranteeing full recognition, is weakly better than c^* at all $p^B \in (0,1)$. If s^* is the high stock offer with participation share $f^{\overline{\theta}_T}(\underline{\theta}_B)$, then deviating towards the low stock offer with participation share $f^{\underline{\theta}_T}(\underline{\theta}_B)$ cannot be profitable, since $f^{\underline{\theta}_T}(\underline{\theta}_B) < f^{\overline{\theta}_T}(\underline{\theta}_B)$ and hence the bidder is fully recognized as a low-type at this new offer as well.

However, it might be that, when $p^B < \pi_S(\underline{\theta}_B)$, the type-$\underline{\theta}_B$ bidder has a profitable deviation from $s^* = (S, f^{\underline{\theta}_T}(\underline{\theta}_B))$ to $(S, f^{\overline{\theta}_T}(\boldsymbol{\mu}^*))$. In order to sustain the equilibrium we thus have to guarantee that

$$(1 - p^B)\Delta(\underline{\theta}_T) \geq (1 - f^{\overline{\theta}_T}(\boldsymbol{\mu}^*))(\mathbb{E}_{p^B}[w(\theta_T)]) - \underline{\theta}_B \tag{4.48}$$

Because

$$(1 - p^B)\Delta(\underline{\theta}_T) > (1 - f^{\overline{\theta}_T}(\underline{\theta}_B))(\underline{\theta}_B + \mathbb{E}_{p^B}[w(\theta_T)]) - \underline{\theta}_B$$

by construction, we can always specify $\boldsymbol{\mu}^*$ so that it retains full recognition at $f^{\underline{\theta}_T}(\underline{\theta}_B)$ while ensuring that (4.48) holds. Notice first how (4.48) requires the equilibrium belief $\boldsymbol{\mu}^*$ *not* to grow too fast in the participation share offered by the bidder: if this were the case, a rather small increase in the offer would be needed for type-$\underline{\theta}_B$ bidder to attract all types of target: in such case, the gains from attracting the high-type targets would surpass the increased costs of a larger participation share. Notice, moreover, that the closer p^B is to

zero, the looser is the belief constraint specified by (4.48), to the point that when $p^B \to 0$ the constraint is actually moot.

Consider now the high-type bidder. When $p^B \leq \pi_S(\overline{\theta}_B)$, the (low) cash offer is always (weakly) optimal. When $p^B > \pi_S(\overline{\theta}_B)$, the high-type bidder could have a profitable deviation towards the high stock offer, as the payoff it expected from the high stock offer exceeds that expected by the low-type bidder. Under equilibrium beliefs $\boldsymbol{\mu}^*$, the payoff expected by the high-type bidder which is mistaken for a low-type bidder when proposing the high stock offer (1) is smaller than the payoff $\overline{\theta}_B$ would obtain by proposing the high stock offer under full recognition and (2) it increases in p^B at the same rate as the payoff expected by $\underline{\theta}_B$ when proposing the same offer. Thus its slope is less than the slope of $\overline{\theta}_B$'s full-recognition payoff from the high stock offer. It follows that either (1) there is no value of p^B such that this payoff is larger than the payoff from the (equilibrium) cash offer or (2) there exists an interval $\left(p^B_-, p^B_+\right) \subseteq (0,1)$ such that $\pi_C \in \left(p^B_-, p^B_+\right)$ and the high-type bidder prefers to deviate to $(S, f^{\overline{\theta}_T}(\underline{\theta}_B))$ than sticking to its equilibrium cash offer if and only if $p^B \in \left(p^B_-, p^B_+\right)$.

Moreover, because of the linearity of payoffs, a necessary and sufficient condition for this *not* to happen is that the payoff from the deviation is not larger than the equilibrium payoff at π^C (which is the value of p^B where the equilibrium payoff is smallest). This is

$$\mathbb{E}_{\pi_C}[w(\theta_T)] - \overline{\theta}_T \geq (1 - f^{\overline{\theta}_T}(\underline{\theta}_B)) \left(\overline{\theta}_B + \mathbb{E}_{\pi_C}[w(\theta_T)]\right) - \overline{\theta}_B \tag{4.49}$$

Using the expression for $f^{\overline{\theta}_T}(\underline{\theta}_B)$ in (4.47) and the one for π_C from (4.18) one obtains condition (4.29). All separating equilibria pass the Intuitive Criterion Test of Definition (9), even though not all of them pass the Equilibrium Dominance Test of Definition (8). Let indeed $p^B \leq \pi_S(\underline{\theta}_B)$, so that $\mathbf{m}^*(\overline{\theta}_B) = (C, \underline{\theta})$ and $\mathbf{m}^*(\underline{\theta}_B) = (S, f^{\underline{\theta}_T}(\underline{\theta}_B))$. Suppose that the out-of-equilibrium message (S, f), with $f > f^{\underline{\theta}_T}(\underline{\theta}_B)$, is observed by the target. If $f < f^{\overline{\theta}_T}(\overline{\theta}_B)$ and the bidder expects all types of target to accept (S, f), then such a message might be rationalized as being sent by $\overline{\theta}_B$ with positive probability. On the other hand, if $f \in (f^{\overline{\theta}_T}(\overline{\theta}_B), f^{\overline{\theta}_T}(\underline{\theta}_B))$, this message could only come from the type-$\underline{\theta}_B$ bidder. Equilibrium dominance then requires that, for any such f, $\boldsymbol{\mu}^*(\overline{\theta}_B \mid (S, f)) = 0$. But because we need $\boldsymbol{\mu}^*(\overline{\theta}_B \mid (S, \cdot))$ to be increasing, this implies

$$\forall f \leq f^{\overline{\theta}_T}(\underline{\theta}_B) \quad , \quad \boldsymbol{\mu}^*(\overline{\theta}_B \mid (S, f)) = 0 \tag{4.50}$$

The restriction thus provided by the application of the EDT, though, does not tamper equilibrium prescriptions: if $\boldsymbol{\mu}^*$ satisfies (4.50), the incentives to deviate by any type of bidder will be weakened. As equilibrium dominance imposes no restriction on separating equilibria when $p^B > \pi_S(\underline{\theta}_B)$, all such equilibria survive the Intuitive Criterion Test. □

We prove here Proposition (5) describing pooling equilibria in which both types of bidder propose a cash offer. We begin by showing that there cannot be such an equilibrium if $p^B > \pi_S(\underline{\theta}_B)$.

Lemma 13. *If $p^B > \pi_S(\underline{\theta}_B)$, there is no PBE such that $\mathbf{m}^*(\overline{\theta}_B) = \mathbf{m}^*(\underline{\theta}_B) =: \mathbf{m}^*$ and $\mathbf{m}^* \in C \times \mathbb{R}_+$.*

Proof. Suppose such an equilibrium exists and let $\boldsymbol{\mu}^*$ be the equilibrium beliefs of the target. As $f^{\overline{\theta}_T}(\boldsymbol{\mu}^*) \leq f^{\overline{\theta}_T}(\underline{\theta}_B)$, the type-$\underline{\theta}_B$ bidder finds the high stock offer more convenient under $\boldsymbol{\mu}^*$ than under any updating rule where it is fully recognized. Because the full recognition payoff of the high stock offer is larger, for all $p^B > \pi_S(\underline{\theta}_B)$ than the payoff from the best cash offer, it follows that the type-$\underline{\theta}_B$ bidder always has an incentive to deviate towards the former. Thus this equilibrium cannot exist. □

Next we show the existence, for all $p^S \leq \pi_S(\underline{\theta}_B)$, of a PBE with both types pooling on the (low) cash offer, thus proving Proposition (5).

Proof of Proposition (5). Let $p^B \leq \pi_S(\overline{\theta}_B)$. Consider the type-$\underline{\theta}_B$ bidder. In order for it not to deviate from the low cash offer to the low stock offer, it is necessary that it finds the two alternatives indifferent, that is, we need to require that it is fully recognized at the low stock offer. This imposes the restriction

$$\forall f \leq f^{\underline{\theta}_T}(\underline{\theta}_B) \quad , \quad \boldsymbol{\mu}^*(\overline{\theta}_B \mid (S, f)) = 0 \tag{4.51}$$

Condition (4.48) must also hold in order for the low-type bidder not to wish to deviate to a high stock offer. Under (4.51) and (4.48) the high-type bidder has no incentive to deviate to any stock offer. Thus the equilibrium is sustained.

As in the proof of Proposition (4), the Equilibrium Dominance Test requires that full recognition of the low-type bidder extends to all $f \leq f^{\overline{\theta}_T}(\underline{\theta}_B)$. This has no effect on the sustainability of the equilibrium action profile. □

We complete the analysis of the equilibria of our model by describing those in which the two types of bidder pool on stock offer.

Lemma 14. *There is no PBE such that $\mathbf{m}^*(\overline{\theta}_B) = \mathbf{m}^*(\underline{\theta}_B) = (S, f^{\underline{\theta}_T}(\boldsymbol{\mu}^*))$.*

Proof. Consider an equilibrium in which both types of bidder pool on a stock offer (S, f^*). By Bayes' rule, $\boldsymbol{\mu}^*(\overline{\theta}_T \mid (S, f^*)) = p^T \in (0, 1)$: since the equilibrium message has no informative content, the target does not update its beliefs upon its reception. It follows that, whatever the pool of targets to which the equilibrium stock offer is directed, this

gives the high-type bidder a smaller payoff than the corresponding full-recognition stock offer. In particular, this excludes that such an equilibrium can consist of a low stock offer, since the latter would always turn out to be worse than the low cash offer for the type-$\overline{\theta}_B$ bidder. □

We now characterize PBEs with pooling on the high stock offer.

Proof of Proposition (6). Consider an equilibrium with $\mathbf{m}^*(\overline{\theta}_B) = \mathbf{m}^*(\underline{\theta}_B) = (S, f^*)$. Because of Bayes' rule, $\boldsymbol{\mu}^*(\overline{\theta}_T \mid (S, f^*)) = p^T \in (0,1)$ and the equilibrium payoff to the high-type bidder is less than the payoff from the full-recognition high stock offer. The latter is better than the best cash offer provided $p^B \geq \pi_S(\overline{\theta}_B)$, so that such an equilibrium cannot arise if $p^B < \pi_S(\overline{\theta}_B)$.

Suppose then $p^B \geq \pi_S(\overline{\theta}_B)$. Since this implies $p^B > \pi_S(\underline{\theta}_B)$, the low-type bidder prefers the high stock offer to any cash offer. We need to ensure that the low-type bidder does not prefer the low stock offer to the high stock offer. This could not be the case if this type of bidder were fully recognized but here we cannot exclude that, when receiving the low stock offer, the target gives positive probability to the bidder's type being $\overline{\theta}_B$. Then we need to ensure that

$$(1-f^*)(\mathbb{E}_{p^B}[w(\theta_T)]) - f^*\underline{\theta}_B \geq p^B\underline{\theta}_B + (1-p^B)\left[(1-f^{\underline{\theta}_T}(\boldsymbol{\mu}^*))w(\underline{\theta}_T) - f^{\underline{\theta}_T}(\boldsymbol{\mu}^*)\underline{\theta}_B\right] \quad (4.52)$$

It is easy to see that this equation imposes an upper bound on $\boldsymbol{\mu}^*$ over $[0, f^*]$. As $\boldsymbol{\mu}^*$ is non-decreasing, this bound is ineffective if (4.52) is satisfied at $f^{\overline{\theta}_T}(\tilde{\mu})$ for a $\tilde{\mu}$ such that $\tilde{\mu}(\overline{\theta}_T \mid (S, f^{\overline{\theta}_T}(\tilde{\mu}))) = p^T$.

Consider the type-$\overline{\theta}_T$ bidder. Since $\boldsymbol{\mu}^*$ is non-decreasing, in case the latter proposed the low stock offer $\left(S, f^{\underline{\theta}_T}(\boldsymbol{\mu}^*)\right)$, the target would associate to the bidder's type being $\overline{\theta}_T$ a probability less than p^T; this makes the low stock offer less convenient than the low cash offer. Thus we need only compare the equilibrium payoff of $\overline{\theta}_B$ with the one from the best cash offer. Indeed, the high-type bidder might prefer the best cash offer to the high stock offer, given that it is not fully recognized. Type-$\overline{\theta}_B$'s payoff from the equilibrium message is

$$(1-f^*)\mathbb{E}_{p^B}[w(\theta_T)] - f^*\overline{\theta}_B \quad (4.53)$$

Such payoff is – as usual – linear in p^B and its slope is $(1-f^*)W$. As $f^* \geq f^{\overline{\theta}_T}(\overline{\theta}_B)$, its slope is less than the one in the full-recognition case. By definition, the full-recognition payoff is larger than the payoff from the best cash offer at all $p^B \in [\pi_S(\overline{\theta}_B), 1)$. Then (S, f^*) will be preferred to the best cash offer on an interval $[p^B_{--}, p^B_{++}] \subset [\pi_S(\overline{\theta}_B), 1)$ provided, similarly to (4.49), that

$$\mathbb{E}_{\pi_C}[w(\theta_T)] - \overline{\theta}_T \leq (1-f^*)\mathbb{E}_{\pi_C}[w(\theta_T)] - f^*\overline{\theta}_B \quad (4.54)$$

Using (4.30) for f^* and (4.18) for π_C, one obtains condition (4.31). Finally, as in the previous cases, we see that an out-of-equilibrium offer (S, f) with $f \in (f^{\underline{\theta}_T}(\underline{\theta}_B), f^{\overline{\theta}_T}(\underline{\theta}_B))$ can only come from a low-type bidder. Equilibrium Dominance thus restricts beliefs so that

$$\forall f < f^{\overline{\theta}_T}(\overline{\theta}_B) \quad , \quad \boldsymbol{\mu}^*(\overline{\theta}_B \mid (S, f)) = 0 \tag{4.55}$$

□

Finally, we show that if a separating equilibrium fails to exist at some $p^B \in (0, 1)$, then a PBE with pooling on stock exists at that p^B.

Proof of Corollary (7). Suppose that p^B is such that the game G does not admit a separating equilibrium. It must be that the high-type bidder prefers the high stock offer to the best cash offer under full recognition as a low-type bidder. Because the participation share of equivalent (in terms of the pool of target to which they are directed) stock offer is decreasing in the probability assigned by the target to the bidder's type being $\overline{\theta}_B$, the high-type bidder prefers, at p^B, the high stock offer f^* to the best cash offer, provided $\boldsymbol{\mu}^* \left(\overline{\theta}_B \mid (S, f^*)\right) = p^T$. Thus the pooling equilibrium is sustained. □

4.5.3 Measuring firm opacity

A potential concern for our analysis is related to the measurement of firm opacity. Our methodology replicates that of Bharath et al. (2009), who form an index on the basis of several measures of adverse selection risk from market microstructure to study the impact of a firm's private information on capital structure decisions. This section first describes in detail the constituents of our index, how it is constructed, and its main properties. Then, it presents some robustness test to validate its use in our empirical analysis.

Our index constituents include (i) the illiquidity measure of Amihud (2002), (ii) the volume − return autocorrelation of Llorente et al. (2002), (iii) the probability of informed trading of Easley et al. (1996), (iv) the adverse selection component of the proportional effective spread of Roll (1984), (v) the reversal coefficient of Pastor and Stambaugh (2003), and (vi) the Amivest liquidity ratio of Cooper et al. (1985) and Amihud et al. (1997). We estimate these measures for all firms i with price and volume data available from the CRSP in any given year y from 1985 to 2014.

Amihud (2002) illiquidity measure is a market microstructure indicator that is interpreted as representative of the price impact, which is increasing in firm opacity. Price impact, in fact, is a measure designed by Kyle (1985) to capture the permanent component of price

change due to trades that move a stock price toward its unobserved fundamental value. Price impact is then higher for firms whose informational gap is larger (i.e., opaque firms), since relatively more information is revealed from trades. Amihud's illiquidity measure, $ILL_{i,y}$, is computed for all firms in our sample as the daily ratio of the absolute value of the stock return to its dollar volume, averaged over all observations in the year.[34]

The return − volume coefficient of Llorente et al. (2002) exploits instead the link between volume-return dynamics and speculation. Following their methodology, for each firm in our sample we estimate the relative importance of information in determining stock return dynamics as the coefficient $c_{2,i,y}$ in the time series regression:

$$r_{i,y,d} = c_{0,i,y} + c_{1,i,y} r_{i,y,d-1} + c_{2,i,y} T_{i,y,d-1} r_{i,y,d-1} + \epsilon_{i,y,d}$$

over all daily observations in a year, where r are daily returns and T is the logarithm of daily turnover (detrended with respect to its mean over the previous 100 observations). The higher the estimated coefficient, the more any stock price change is driven by information and then the more opaque the firm is.

The probability of informed trading of Easley et al. (1996) is an assessment of the likelihood of an informed order. It results from imbalances in the order flow: in principle, in fact, uninformed orders to buy and sell a firm stock occur randomly and therefore imbalances signal informed trading. Then, orders for opaque firms are more clustered and the probability of informed trading for opaque firms is higher. We obtain $PIN_{i,y}$ for firms with stock traded on the NYSE or AMEX between 1985 and 2001 from Easley et al. (2010).

The adverse selection component of the proportional effective spread of Roll (1984) exploits return autocorrelation to quantify the informational nature of price dynamics. Uninformed trading is associated with the negative autocorrelation of returns, since a variation in stock price is not accompanied by a change in the market expectation of its fundamental value. On the contrary, informed trades determine the positive autocorrelation of returns as the market gradually updates its expectation of a stock's fundamental value. We then estimate the adverse selection component of the proportional effective spread of a firm's stock, filtering its realized returns with a measure of its time-varying expected return according to George et al. (1991). In particular, $RAD_{i,y}$ is computed as $1 - \pi^2_{1,i,y}$ form the regression:

$$FRS_{i,y,d} = \pi_{0,i,y} + \pi_{1,i,y} RS_{i,y,d} + \epsilon i, y, d$$

[34] Amihud (2002) shows that this measure is strongly positively related to intra-day estimates of price impact. As suggested by Amihud (2002), we rescale the values by multiplying by 10^6 and, as suggested by Hasbrouck (2009), use a square root transformation.

over all daily observations in a year, where $RS_{i,y}$ is the proportional effective spread of Roll (1984) calculated on the basis of 60-day rolling autocovariances of returns as

$$RS_{i,y,d} = 200\sqrt{-cov(r_{i,y,d}, r_{i,y,d-1})} \qquad \text{if} \qquad cov(r_{i,y,d}, r_{i,y,d-1}) < 0$$

$$RS_{i,y,d} = 200\sqrt{cov(r_{i,y,d}, r_{i,y,d-1})} \qquad \text{otherwise}$$

and $FRS_{i,y}$ is the filtered proportional effective spread, computed as $RS_{i,y}$ but on the basis of the autocovariances of the residuals from a regression of daily returns on their expected return series (estimated with a market model over observations of the previous year). More opaque firms are characterized by a larger fraction of the proportional effective spread due to adverse selection.

The reversal coefficient of Pastor and Stambaugh (2003) results from the interaction between a stock's return and its lagged order flow. In particular, the intuition is that the greater is the extent of a firm's private information, the lower its stock liquidity and the higher the estimated return reversal for a given dollar volume. Following their methodology, for each firm in our sample we estimate $GAM_{i,y}$ as the coefficient $\gamma_{i,y}$ of the one-period-lagged signed volume in the time series regression of daily excess returns:[35]

$$r^e_{i,y,d} = \theta_{i,y} + \varphi_{i,y} r_{i,y,d-1} + \gamma_{i,y} V_{i,y,d-1} + \epsilon_{i,y,d}$$

over all daily observations in a year, where $V_{i,y}$ is daily dollar volume signed according to the contemporaneous excess return. The higher the estimated coefficient, the more opaque the firm.

Finally, the Amivest liquidity ratio of Cooper et al. (1985) and Amihud et al. (1997) is used to capture the fact that liquidity mitigates the price impact of large volumes. It is computed for all firms in our sample as the square root of the ratio of a firm's stock daily dollar volume to its absolute return, averaged over all daily observations in a year and preceded by a negative sign. The higher its value, the higher the opacity of the firm.

Table 4.5 presents summary statistics for all our index constituents and Spearman's rank correlations among their standardized values for all firms with data available on CRSP in the period between 1985 and 2014. Our estimates are similar to those of Bharath et al. (2009) in a partially overlapping subsample.

Although all the proposed measures are linked to firm opacity, information is not their only driver. We then isolate the common informational element by estimating the first

[35] Excess returns are with respect to the value-weighted market return of all firms on CRSP in the corresponding period.

Table 4.5: Descriptive statistics of index constituents.

This table reports summary statistics for the constituents of our index of firm opacity. ILL is Amihud (2002) illiquidity measure, C2 is the volume-return autocorrelation of Llorente et al. (2002), PIN is the probability of informed trading of Easley et al. (1996), RAD is the adverse selection component of the proportional effective spread of Roll (1984), GAM is the reversal coefficient of Pastor and Stambaugh (2003), and LR is the Amivest liquidity ratio of Cooper et al. (1985) and Amihud et al. (1997). Panel a. presents cross-sectional statistics over the sample period 1985-2014. Panel b. reports the Spearman's rank correlations among the standardized values of the index constituents. The superscript ^{a} denotes statistical significance at the 1% level.

Panel a.	Obs.	Mean	Median	St. Dev.
ILL	146,882	1.01	0.37	1.68
C2	118,084	0.02	0.02	0.23
PIN	31,103	0.21	0.20	0.08
RAD	146,932	0.37	0.45	0.53
GAM	146,882	0.85	0.04	4.19
LR	146,853	-11.53	-3.09	23.09

Panel b.	ILL	C2	PIN	RAD	GAM	LR
ILL	1					
C2	0.0723^a	1				
PIN	0.6955^a	0.0111	1			
RAD	-0.2431^a	-0.0277^a	-0.1548^a	1		
GAM	0.8227^a	0.0565^a	0.5457^a	-0.1716^a	1	
LR	0.9908^a	0.0875^a	0.6848^a	-0.2419^a	0.8247^a	1

principal component of the correlation matrix of our standardized index constituents in each year. On average, 40% of cross-sectional variance is accounted for by the first principal component and in most years only the first eigenvalue is larger than one. Moreover, the elements of the first eigenvector are mostly positive and their magnitude is stable over time, confirming that each constituent adds positively to the index.

We form the index of firm opacity by combining standardized index constituents according to the corresponding contemporaneous loadings on the first principal component. According to our index, the opacity of firm i in year y is computed on the basis of our six index constituents x, standardized across all firms in the given year, as

$$Index_{i,y} = \sum_{j=1}^{6} w_{j,y}\bar{x}_{i,y} \qquad \text{where} \qquad w_{j,y} = PC(\bar{x}_{i,y})$$

Higher values of the index are associated with higher opacity for the specific firm in the given year. In each year, as well as overall, the mean index value is zero by construction, the median is slightly negative, and the standard deviation is 1.42.

The literature has linked firm opacity to several firm characteristics. In Table 4.6 we investigate, for all firms with data available on Compustat, the distribution of these information-related characteristics across different classes of opacity formed on the basis of our index. These variables include firm size, capital expenditures, R&D expenses, cash holdings and leverage.

Table 4.6: Firm characteristics across levels of firm opacity.

This table reports means for alternative firm characteristics across levels of opacity. Each firm is classified each year on the basis of its index of opacity. Assets and Sales are in millions of dollars and adjusted for inflation; Capex and R&D are, respectively, capital expenditures and R&D expenses, both over total assets; Cash and Lev are, respectively cash holdings and financial debt, both over total assets. Capex, R&D, Cash and Lev are expressed in % terms. Only firms with fiscal year ending in December are considered.

Opacity	Assets	Sales	Capex	R&D	Cash	Lev
Lowest	17,039.28	6,316.26	6.51	5.81	8.88	23.20
2	1,587.28	803.19	6.39	9.43	12.35	21.16
3	676.81	352.27	6.23	11.71	13.00	22.24
4	542.59	175.64	6.01	13.41	12.26	23.04
Highest	164.40	73.75	4.93	12.43	10.74	27.08

Not surprisingly, more opaque firms are, on average, smaller, in terms of both total assets and sales. Size follows a steadily decreasing trend as opacity increases. We then observe fewer capital expenditures (Capex) as opacity grows. Interestingly, we find that more opaque firms report, on average, higher levels of R&D expenses (R&D). This evidence is consistent with more innovative firms being inevitably more opaque due to the uncertainty in their future prospects. All these trends support our claim that our index of firm opacity captures the informational dimension at the core of our analysis. Finally we observe that the most opaque firms are on average more leveraged and hoard more cash, consistent with the pecking-order theory of financing.

4.5.4 Additional tables and figures

Table 4.7: Variables Definition.

This table summarizes the variables used in our empirical analysis, with a brief description and their sources.

Variable	*Definition*	*Source*
Panel a. Dependent variables.		
MP	Method of payment: a dummy variable equal to 1 for stock offers and 0 otherwise.	SDC Thomson
PRM	Premium: the percentage amount by which the offering price exceeds target's stock price four weeks before the announcement.	SDC Thomson
Panel b. Firm opacity.		
Bidder Opacity Target Opacity	Index: measured on the basis of the common cross-sectional variation of (i) the illiquidity measure of Amihud (2002), (ii) the volume − return autocorrelation of Llorente et al. (2002), (iii) the probability of informed trading of Easley et al. (1996), (iv) the adverse selection component of the proportional effective spread of Roll (1984), (v) the reversal coefficient of Pastor and Stambaugh (2003), and (vi) the Amivest liquidity ratio of Cooper et al. (1985) and Amihud et al. (1997). The index is formed annually on the basis of the first principal component of the standardized values of these measures (with respect to all firms in CRSP).[36]	CRSP
Panel c. Controls: deal characteristics.		
Deal Value	Total value of consideration paid by the acquiror, excluding fees and expenses.	SDC Thomson
Deal Materiality	The ratio between the undisturbed market capitalization of the target over the sum of the undisturbed market capitalizations of the bidder and the target 63 days preceding the bid.	CRSP

[36]We also compute the index of opacity under alternative standardizations of its components, at the industry level (according to the Fama and French 48-industry classification) or with respect to size. The results of the empirical analysis employing these alternative measures are available upon request.

Run-up	The cumulative return of the target's stock price in the window [-62,-1] with respect to the announcement date.	CRSP and SDC Thomson
Synergies	The market capitalization-weighted average of the bidder's and target's cumulative abnormal returns in the window [-62, 126], as in Bradley et al. (1988)	CRSP and SDC Thomson
Tender Offer	A dummy variable that equals 1 if the bid is reported as a tender offer, 0 otherwise.	SDC Thomson
Diversifying	A dummy variable that equals 1 if the deal involves the bidder and target operating in different two-digit SIC codes, 0 otherwise.	SDC Thomson
Friendly	A dummy variable that equals 1 if the deal is classified as friendly, 0 otherwise.	SDC Thomson
Toehold	A dummy variable that equals 1 if the bidder owns an interest in excess of 5% (threshold for which a bidder has to file a Schedule 13D with the SEC) in the target pre-bid.	SDC Thomson
Hostile	A dummy variable that equals 1 if the deal is classified as hostile, 0 otherwise.	SDC Thomson
Poison Pill	A dummy variable that equals 1 if the target has a poison pill, 0 otherwise.	SDC Thomson
Rumored	A dummy variable that equals 1 if the bid is anticipated by some leakage of information before the announcement according to SDC Thomson One Banker, 0 otherwise.	SDC Thomson
Rivaled	A dummy variable that equals 1 if the bid is challenged by a rival bid, 0 otherwise.	SDC Thomson
Terminated	A dummy variable that equals 1 if the deal has been terminated by the target, 0 otherwise	SDC Thomson
Withdrawn	A dummy variable that equals 1 if the deal has been withdrawn by the bidder, 0 otherwise	SDC Thomson
Panel d. Controls: bidder and target firm characteristics.		
Size	The logarithmic transformation of the total assets of the bidder or the target.	Compustat
Leverage	The bidder's short and long term debt over total assets.	Compustat
Cash Holdings	The bidder's cash holdings over total assets.	Compustat
Cash Flows	The bidder's operating cash flows over total assets.	Compustat

Market-to-Book	The market capitalization over book value of equity of the bidder and the target.	Compustat
Invest. Opp.	Investment opportunities: bidder's capital expenditures and R&D expenses over total assets.	Compustat
Panel e. Controls: institutional and macro environment.		
Capital Gain	A dummy variable to identify bids announced from 1989 to 1996, a period of good market performance and a high (28%) tax rate on capital gains.	
Wave	A dummy variable that equals 1 if the deal occurs in a period of exceptional concentration of merger activity, as for Harford (2005), in the industry of either the bidder, the target, or both, and 0 otherwise.	SDC Thomson
Uncertainty	The VIX index.	CBOE
Sentiment	Robert Shiller's cyclically adjusted P/E index.	Prof. R. Shiller
Liquidity	The spread between the average interest rate on Commercial and Industrial Loans and the Fed Funds Rate.	FRED
Credit Spread	The yield spread between 20-year Baa and Aaa corporate bonds	FRED
Term Spread	The yield spread between the 10-year government bond and the 3-month T-Bill	FRED

Table 4.8: Descriptive Statistics: bid characteristics.

This table reports summary statistics for the bids included in the sample. Means and standard deviations are computed across the entire sample and conditional on the method of payment. In the last column, the result of a parametric t-test of the equivalence of means across methods of payment is presented. The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	All Bids Mean/Sd	Cash Mean/Sd	Stock Mean/Sd	Cash-Stock Diff./t-stat
Deal Value ($ mil.)	1,589	1,080	1,998	-918***
	6,304	2,110	8,225	(-4.26)
Deal Materiality (%)	18.9	15.8	21.9	-6.1***
	20.1	19.5	20.3	(-6.70)
Premium (%)	43.6	49.1	39.2	9.9***
	38.7	38.4	38.4	(6.44)
Run-up (%)	15.8	17.1	14.7	2.4*
	35.4	33.0	37.3	(1.75)
Synergies (%)	4.4	6.9	2.7	3.6
	57.1	37.8	70.2	(1.46)
Tender Offer (%)	19.6	44.3	1.2	44.1***
	39.7	49.8	10.9	(31.51)
Hostile (%)	4.9	9.5	1.6	7.8***
	21.6	29.3	12.7	(9.09)
Friendly (%)	91.8	85.0	96.6	-11.6***
	27.5	35.7	18.1	(-10.80)
Diversifying (%)	35.3	44.0	29.1	14.9***
	47.8	49.7	45.4	(8.59)
Toehold (%)	4.1	7.3	1.6	5.4***
	19.8	26.0	13.5	(6.89)
Poison Pill (%)	2.1	4.5	0.3	4.2***
	14.2	20.8	5.7	(7.10)
Rumored (%)	3.9	4.4	3.6	0.8
	19.5	20.6	18.6	(1.16)
Rivaled (%)	8.7	14.8	4.4	10.5***
	28.2	35.5	20.4	(9.57)
Withdrawn (%)	15.8	17.1	14.9	2.2*
	36.5	37.7	35.6	(1.66)
Wave (%)	30.3	20.0	31.2	-2.1
	40.0	45.4	46.3	(-1.29)
Number of observations	3,141	1,309	1,832	-

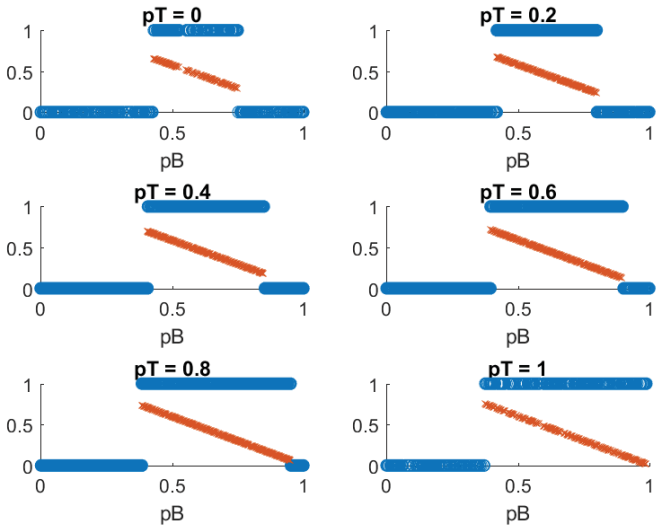


Figure 4.13: Existence of a stock pooling equilibrium (blue dots) and associated bid premium (orange dots) over p^B.

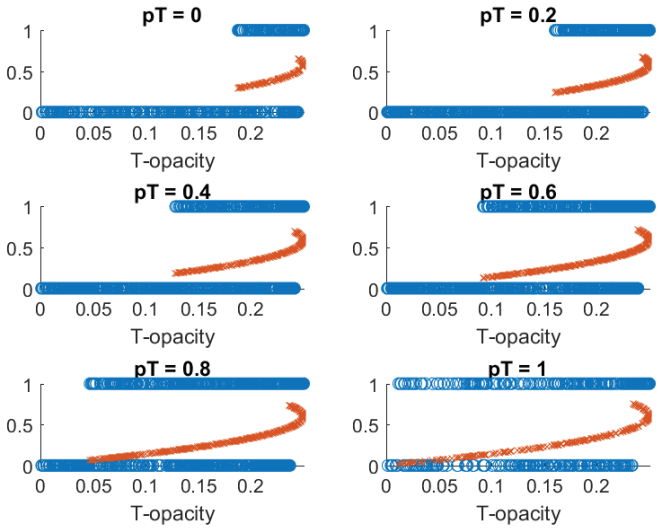


Figure 4.14: Existence of a stock pooling equilibrium and associated bid premium over ω_T.

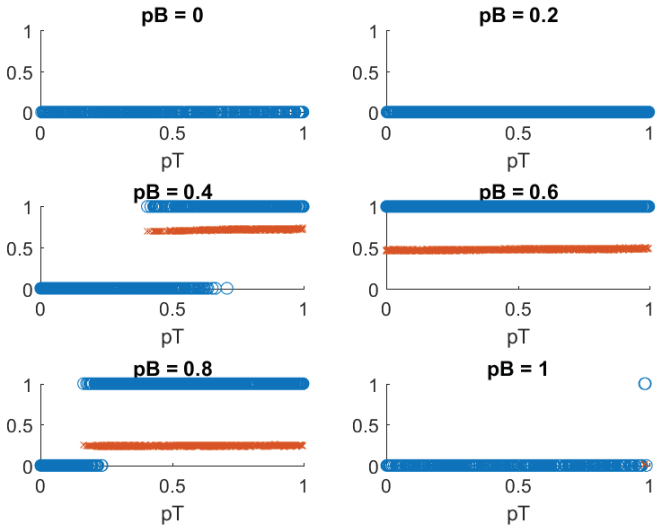


Figure 4.15: Existence of a stock pooling equilibrium and associated bid premium over p^T.

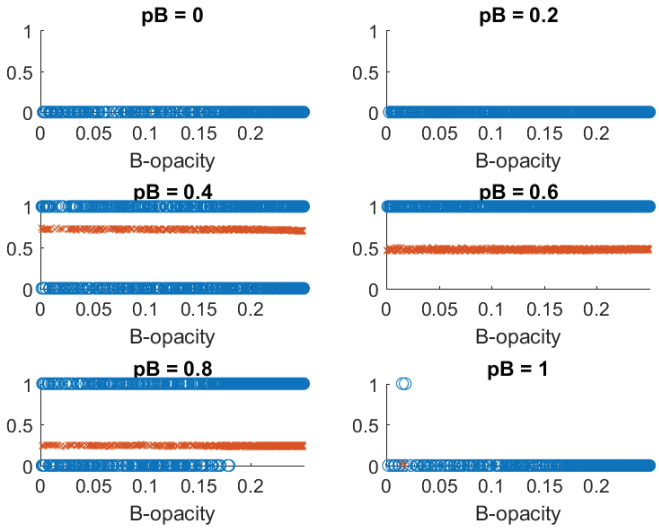


Figure 4.16: Existence of a stock pooling equilibrium and associated bid premium over ω_B.

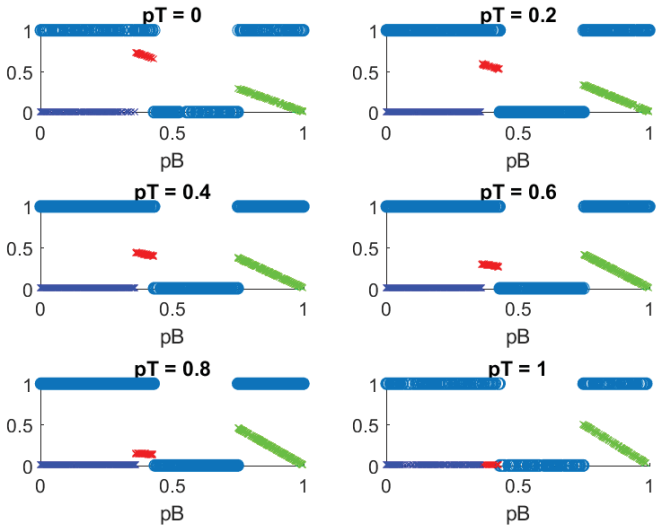


Figure 4.17: Existence of a separating equilibrium (blue dots) and associated bid premium (blue dots for type-A separating equilibria; red dots for type-B separating equilibria; green dots for type-C separating equilibria) over p^B.

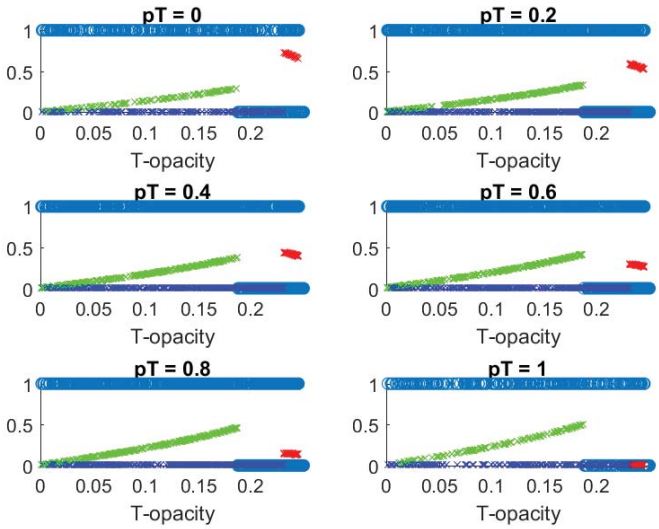


Figure 4.18: Existence of a separating equilibrium and associated bid premium over ω_T.

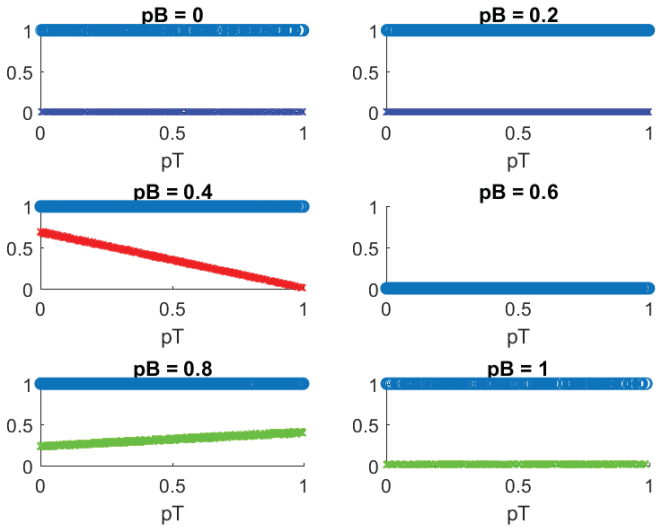


Figure 4.19: Existence of a separating equilibrium and associated bid premium over p^T.

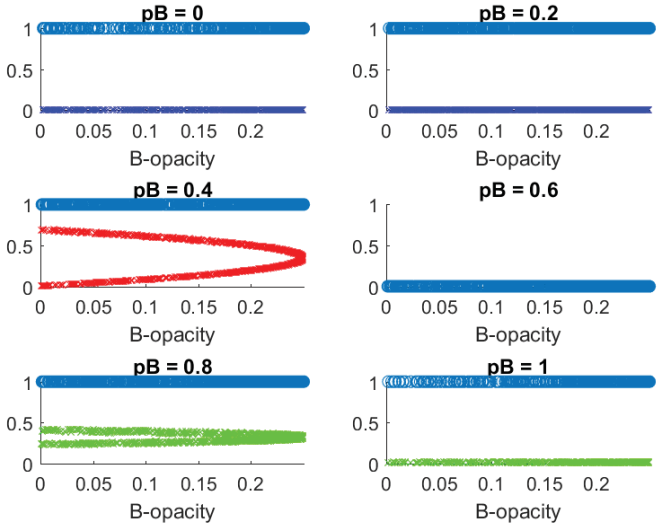


Figure 4.20: Existence of a separating equilibrium and associated bid premium over ω_B.

Bibliography

Accetturo, A., Bugamelli, M. and Lamorgese, A. (2017), 'Law enforcement and political participation: Italy, 1861-1865', *Journal of Economic Behavior & Organization* **104**, 224–245.

Acemoglu, D., Cantoni, D., Johnson, S. and Robinson, J. A. (2011), 'The consequences of radical reform: the French Revolution', *American Economic Review* **101**(7), 3286–3307.

Acemoglu, D. and Jackson, M. O. (2017), "Social norms and the enforcement of laws", *Journal of the European Economic Association* **15**(2), 245–295.

Acemoglu, D., Johnson, S. and Robinson, J. A. (2005), 'Institutions as a fundamental cause of long-run growth', *in* P. Aghion and S. N. Durlauf, eds, 'Handbook of Economic Growth, Volume 1A', Elsevier, San Diego.

Acemoglu, D. and Robinson, J. A. (2012), *Why nations fail: the origins of power, prosperity, and poverty*, Random House, New York.

Acemoglu, D. and Verdier, T. (1998), 'Property rights, corruption and the allocation of talent: a general equilibrium approach', *Economic Journal* **108**, 1381–1403.

Al-Najjar, N. I. (2004), 'Aggregation and the law of large numbers in large economies', *Games and Economic Behavior* **47**(1), 1–35.

Aldashev, G., Chaara, I., Platteau, J.-P. and Wahhaj, Z. (2012), 'Using the law to change the custom', *Journal of Development Economics* **97**(2), 182–200.

Alesina, A. and Giuliano, P. (2015), 'Culture and institutions', *Journal of Economic Literature* **53**(4), 898–944.

Alesina, A., Giuliano, P. and Nunn, N. (2013), 'On the origins of gender roles: women and the plough', *Quarterly Journal of Economics* **128**(2), 469–530.

Alesina, A. and Reich, B. (2015), 'Nation building'. NBER Working Paper no. 18839.

Alesina, A., Tabellini, G. and Trebbi, F. (2017), 'Is Europe an optimal political area?'. NBER Working Paper no. 23325.

Allen, R. C. (2001), 'The great divergence in European wages and prices from the Middle Ages to the First World War', *Explorations in Economic History* **38**(4), 411–447.

Amihud, Y. (2002), 'Illiquidity and stock returns: cross-section and time-series effects', *Journal of Financial Markets* **5**(1), 31–56.

Amihud, Y., Lev, B. and Travlos, N. (1990), 'Corporate control and the choice of investment financing: the case of corporate acquisitions', *The Journal of Finance* **45**(2), 603–616.

Amihud, Y., Mendelson, H. and Lauterbach, B. (1997), 'Market microstructure and securities values: evidence from the Tel Aviv Stock Exchange', *Journal of Financial Economics* **45**(3), 365–390.

Amodio, F. (2012), 'Hard to forget: long-lasting effects of social capital accumulation shocks'. AICCON Working Paper no. 105.

Antoniou, A., Arbour, P. and Zhao, H. (2008), 'How much is too much: are merger premiums too high?', *European Financial Management* **14**(2), 268–287.

Banerjee, R., Baul, T. and Rosenblat, T. (2015), 'On self selection of the corrupt into the public sector', *Economics Letters* **127**, 43–46.

Banfield, E. C. (1958), *The moral basis of a backward society*, Free Press, New York.

Barfort, S., Harmon, N., Hjorth, F. and Olsen, A. L. (2017), 'Sustaining honesty in public service: the role of selection'. Working paper.

Barone, G. and Mocetti, S. (2014), 'Natural disasters, growth and institutions: a tale of two earthquakes', *Journal of Urban Economics* **84**, 52–66.

Battigalli, P. (2006), 'Rationalization in signaling games: theory and applications', *International Game Theory Review* **8**(1), 67–93.

Battigalli, P. and Siniscalchi, M. (2002), 'Strong belief and forward induction reasoning', *The Journal of Economic Theory* **106**(2), 356–391.

Battigalli, P. and Siniscalchi, M. (2003), 'Rationalization and incomplete information', *Advances in Theoretical Economics* **3**(1), Article 3.

Beatty, R. and Ritter, J. (1986), 'Investment banking, reputation, and the underpricing of initial public offerings', *Journal of Financial Economics* **15**(1), 213–232.

Ben Amos, I. K. (1991), 'Failure to become freemen: urban apprentices in early modern England', *Social History* **16**(2), 155–172.

Ben Zeev, N., Mokyr, J. and van der Beek, M. (2017), 'Flexible supply of apprenticeship in the British industrial revolution', *The Journal of Economic History* **77**(1), 208–250.

Berger, P. and Ofek, E. (1995), 'Diversification's effect on firm value', *Journal of Financial Economics* **37**(1), 39–65.

Berkovitch, E. and Khanna, N. (1991), 'A theory of acquisition markets: mergers versus tender offers, and golden parachutes', *Review of Financial Studies* **4**(1), 149–174.

Berkowitz, D., Pistor, K. and Richard, J.-F. (2003*a*), 'Economic development, legality, and the transplant effect', *European Economic Review* **47**(1), 165–195.

Berkowitz, D., Pistor, K. and Richard, J.-F. (2003*b*), 'The transplant effect', *American Journal of Comparative Law* **51**(1), 163–203.

Betton, S., Eckbo, B. E., Thompson, R. and Thorburn, K. S. (2014), 'Merger negotiations with stock market feedback', *The Journal of Finance* **69**(4), 1705–1745.

Bharath, S., Pasquariello, P. and Wu, G. (2009), 'Does asymmetric information drive capital structure decisions?', *Review of Financial Studies* **22**(8), 3211–3243.

Bisin, A. and Verdier, T. (2011), 'The economics of cultural transmission and socialization', *in* J. Benhabib, A. Bisin and M. O. Jackson, eds, 'Handbook of Social Economics, Volume 1A', Elsevier, San Diego.

Bisin, A. and Verdier, T. (2017), 'On the joint evolution of culture and institutions'. NBER Working Paper no. 23375.

Bitonti, A. (2012), *Luoghi, lingue, contatto. Italiano, dialetti e francoprovenzale in Puglia*, Congedo Editore, Galatina.

Boone, A. L., Lie, E. and Liu, Y. (2014), 'Time trends and determinants of the method of payment in M&As', *Journal of Corporate Finance* **27**, 296–304.

Boone, A. L. and Mulherin, J. H. (2007), 'How are firms sold?', *The Journal of Finance* **62**(2), 847–875.

Bracco, E., De Paola, M. and Green, C. P. (2015), 'Long lasting differences in civic capital: evidence from a unique immigration event in Italy', *Journal of Economic Behavior & Organization* **120**, 160–173.

Bradley, M., Desai, A. and Kim, E. (1988), 'Synergistic gains from corporate acquisitions and their division between the stockholders of target and acquiring firms', *Journal of Financial Economics* **21**(1), 3–40.

Brandolini, A. and Cannari, L. (1994), 'Methodological appendix: the Bank of Italy Survey of household income and wealth', *in* A. Ando, L. Guiso and I. Visco, eds, 'Saving and the accumulation of wealth. Essays on Italian households and government saving behavior', Cambridge University Press.

Brollo, F., Nannicini, T., Perotti, R. and Tabellini, G. (2013), 'The political resource curse', *American Economic Review* **103**(5), 1759–1796.

Brown, D. and Ryngaert, M. (1991), 'The mode of acquisition in takeovers: taxes and asymmetric information', *The Journal of Finance* **46**(2), 653–669.

Camussi, S. and Mancini, A. L. (2016), 'Individual trust: does quality of public services matter?'. Bank of Italy, Working Paper no. 1069.

Cavalli Sforza, L. L. and Feldman, M. W. (1981), *Cultural transmission and evolution: a quantitative approach*, Princeton University Press.

Chae, J. (2005), 'Trading volume, information asymmetry, and timing information', *The Journal of Finance* **60**(1), 413–442.

Charron, N., Lapuente, V. and Dijkstra, L. (2014), 'Regional governance matters: quality of government within European Union member states', *Regional Studies* **48**(1), 68–90.

Chatterjee, S., John, K. and Yan, A. (2012), 'Takeovers and divergence of investor opinion', *Review of Financial Studies* **25**(1), 227–277.

Chemmanur, T., Paeglis, I. and Simonyan, K. (2009), 'The medium of exchange in acquisitions: does private information of both acquirer and target matter?', *Journal of Corporate Finance* **15**(5), 523–542.

Chen, J., Hong, H. and Stein, J. (2002), 'Breadth of ownership and stock returns', *Journal of Financial Economics* **66**(2), 171–205.

Cheng, P., Li, J. and Tong, W. (2016), 'Target information asymmetry and acquisition price', *Journal of Business Finance & Accounting* **43**(7-8), 976–1016.

Chidambaran, N., John, K., Shangguan, Z. and Vasudevan, G. (2010), 'Hot and cold merger markets', *Review of Quantitative Finance and Accounting* **34**(3), 327–349.

Cho, I. K. and Kreps, D. M. (1987), 'Signaling games and stable equilibria', *Quarterly Journal of Economics* **102**(2), 179–221.

Ciocca, P. (2013), 'Brigantaggio ed economia nel Mezzogiorno d'Italia, 1860-1870', *Rivista di Storia Economica* **29**(1), 3–30.

Clark, G. (2014), *The son also rises: surnames and the history of social mobility*, Princeton University Press.

Clausen, B., Kraay, A. and Nyiri, Z. (2011), 'Corruption and confidence in public institutions: evidence from a global survey', *World Bank Economic Review* **25**(2), 212–249.

Comment, R. and Schwert, G. (1995), 'Poison or placebo? Evidence on the deterrence and wealth effects of modern antitakeover measures', *Journal of Financial Economics* **39**(1), 3–43.

Cooper, K. S., Groth, J. C. and Avera, W. E. (1985), 'Liquidity, exchange listing, and common stock performance', *Journal of Economics and Business* **37**(1), 19–33.

Cording, M., Christmann, P. and Bourgeois Iii, L. (2002), 'A focus on resources in M&A success: a literature review and research agenda to resolve two paradoxes', *Academy of Management Proceeds* .

Corona, G. (1995), *Demani ed individualismo agrario nel Regno di Napoli*, ESI, Roma.

Cowley, E. and Smith, S. (2014), 'Motivation and mission in the public sector: evidence from the World Values Survey', *Theory and Decision* **76**(2), 241–263.

De Francesco, M. R. (1988), 'Equilibri territoriali e divisioni demaniali nel Contado di Molise', *in* A. Massafra, ed., 'Il Mezzogiorno pre-unitario. Economia, società e istituzioni', Dedalo, Bari.

De la Croix, D., Doepke, M. and Mokyr, J. (2017), 'Clans, guilds and markets: apprenticeship institutions and growth in the pre-industrial economy', *Quarterly Journal of Economics* (forthcoming).

De Mauro, T. (1963), *Storia linguistica dell'Italia unita*, Laterza, Bari.

De Munck, B., Kaplan, S. L. and Soly, H. (2007), *Learning on the shop floor: historical perspectives on apprenticeship*, Berghahn Books, New York.

De Munck, B. and Soly, H. (2007), '"Learning on the shop floor" in historical perspective', *in* B. De Munck, S. L. Kaplan and H. Soly, eds, 'Learning on the shop floor: historical perspectives on apprenticeship', Berghahn Books, New York.

De Nardi, M. (2004), 'Wealth inequality and intergenerational links', *The Review of Economic Studies* **71**(3), 743–768.

De Salvio, A. (1908), 'Relics of Franco-Provençal in Southern Italy', *Publications of the Modern Language Association of America* **23**(1), 45–79.

Del Monte, A. and Papagni, E. (2001), 'Public expenditure, corruption, and economic growth: the case of Italy', *European Journal of Political Economy* **17**(1), 1–16.

Del Monte, A. and Papagni, E. (2007), 'The determinants of corruption in Italy: regional panel data analysis', *European Journal of Political Economy* **23**(2), 379–396.

Demsetz, H. and Lehn, K. (1985), 'The structure of corporate ownership: causes and consequences', *Journal of Political Economy* **93**(6), 1155–1177.

Di Liberto, A. and Sideri, M. (2015), 'Past dominations, current institutions and the Italian regional economic performance', *European Journal of Political Economy* **38**, 12–41.

Diether, K., Malloy, C. and Scherbina, A. (2002), 'Differences of opinion and the cross section of stock returns', *The Journal of Finance* **57**(5), 2113–2141.

Doepke, M. and Zilibotti, F. (2008), 'Occupational choice and the spirit of capitalism', *Quarterly Journal of Economics* **123**(2), 747–793.

Doepke, M. and Zilibotti, F. (2014), 'Culture, entrepreneurship and growth', *in* P. Aghion and S. N. Durlauf, eds, 'Handbook of Economic Growth, Volume 2A', Elsevier, Oxford.

Doepke, M. and Zilibotti, F. (2017), 'Parenting with style: altruism and paternalism in intergenerational preference transmission', *Econometrica* **85**(5), 1331–1371.

Dunlop, O. J. and Denman, R. D. (1912), *English apprenticeship and child labor*, T. F. Unwin, London.

Earle, P. (1989), *The making of the English middle class: business, society, and family life in London, 1660-1730*, Methuen, London.

Easley, D., Hvidkjaer, S. and O'Hara, M. (2010), 'Factoring information into returns', *Journal of Financial and Quantitative Analysis* **45**(2), 293.

Easley, D., Kiefer, N. M., O'Hara, M. and Paperman, J. B. (1996), 'Liquidity, information, and infrequently traded stocks', *The Journal of Finance* **51**(4), 1405–1436.

Easley, D. and O'Hara, M. (2005), 'Information and the cost of capital', *The Journal of Finance* **59**(4), 1553–1583.

Eckbo, B. (2009), 'Bidding strategies and takeover premiums: a review', *Journal of Corporate Finance* **15**(1), 149–178.

Eckbo, E., Giammarino, R. and Heinkel, R. (1990), 'Asymmetric information and the medium of exchange in takeovers: theory and tests', *Review of Financial Studies* **3**(4), 651–675.

Elbaum, B. and Singh, N. (1995), 'The economic rationale of apprenticeship training: some lessons from British and US experience', *Industrial Relations: A Journal of Economy and Society* **34**(4), 593–622.

Epstein, S. R. (1998), 'Craft guilds, apprenticeship, and technological change in preindustrial Europe', *The Journal of Economic History* **58**(3), 684–713.

Epstein, S. R. (2003), 'Apprenticeship', *in* J. Mokyr, ed., 'The Oxford Encyclopedia of Economic History', Oxford University Press.

Epstein, S. R. (2008), 'Craft guilds in the pre-modern economy: a discussion', *The Economic History Review* **61**(1), 155–174.

Epstein, S. R. and Prak, M. (2008), *Guilds, innovation and the European economy, 1400-1800*, Cambridge University Press.

Faccio, M. and Masulis, R. (2005), 'The choice of payment method in European mergers and acquisitions', *The Journal of Finance* **60**(3), 1345–1388.

Fernández, R. (2013), 'Cultural change as learning: the evolution of female labor force participation over a century', *American Economic Review* **103**(1), 472–500.

Ferraz, C. and Finan, F. (2011), 'Electoral accountability and corruption: evidence from the audits of local governments', *American Economic Review* **101**(4), 1274–1311.

Finlay, R. (1981), *Population and metropolis. The demography of London 1580-1650*, Cambridge University Press.

Fishman, M. J. (1989), 'Preemptive bidding and the role of the medium of exchange in acquisitions', *The Journal of Finance* **44**(1), 41–57.

Fisman, R. and Gatti, R. (2012), 'Decentralization and corruption: evidence across countries', *Journal of Public Economics* **83**(3), 325–345.

Fisman, R. and Miguel, E. (2007), 'Corruption, norms, and legal enforcement: evidence from diplomatic parking tickets', *Journal of Public Economics* **115**(6), 1020–1048.

Fisman, R., Paravisini, D. and Vig, V. (2017), 'Cultural proximity and loan outcomes', *American Economic Review* **107**(2), 457–492.

Fogli, A. and Veldkamp, L. (2011), 'Nature or nurture? Learning and the geography of female labor force participation', *Econometrica* **79**(4), 1103–1138.

Franks, J., Harris, R. and Mayer, C. (1988), 'Means of payment in takeovers: results from the United Kingdom and the United States', *in* J. Franks, R. Harris, C. Mayer and A. J. Auerbach, eds, 'Corporate takeovers: causes and consequences', University of Chicago Press.

Fudenberg, D. and Tirole, J. (1991), 'Perfect Bayesian equilibrium and sequential equilibrium', *The Journal of Economic Theory* **53**(2), 236–260.

Fuller, K., Netter, J. and Stegemoller, M. (2002), 'What do returns to acquiring firms tell us? Evidence from firms that make many acquisitions', *The Journal of Finance* **57**(4), 1763–1793.

Galor, O. and Özak, Ö. (2016), 'The agricultural origins of time preference', *American Economic Review* **106**(10), 3064–3103.

George, T. J., Kaul, G. and Nimalendran, M. (1991), 'Estimation of the bid-ask spread and its components: A new approach', *Review of Financial Studies* **4**(4), 623–656.

Gillis, W. E. (2009), 'Acquisition financing: does how you pay for it have implications for success?', *The Academy of Management Perspectives* **23**(4), 96–97.

Giorgiantonio, C., Orlando, T., Palumbo, G. and Rizzica, L. (2016), 'Incentives and selection in public employment'. Bank of Italy, Occasional Paper no. 342.

Golden, M. and Picci, L. (2005), 'Proposal for a new measure of corruption, illustrated with Italian data', *Economics and Politics* **17**(1), 37–75.

Gonzalez, F. M. and Shi, S. (2010), 'An equilibrium theory of learning, search, and wages', *Econometrica* **78**(2), 509–537.

Grassby, R. (1995), *The business community of seventeenth-century England*, Cambridge University Press.

Guinnane, T. W. (1994), 'A failed institutional transplant: Raiffeisen credit cooperatives in Ireland, 1894-1914', *Explorations in Economic History* **31**(1), 38–61.

Guiso, L., Herrera, H. and Morelli, M. (2016), 'Cultural differences and institutional integration', *Journal of International Economics* **99**(S1), S97–S113.

Guiso, L., Sapienza, P. and Zingales, L. (2004), 'The role of social capital in financial development', *American Economic Review* **94**(3), 526–556.

Guiso, L., Sapienza, P. and Zingales, L. (2006), 'Does culture affect economic outcomes?', *Journal of Economic Perspectives* **20**(2), 23–48.

Guiso, L., Sapienza, P. and Zingales, L. (2008), 'Alfred Marshall lecture on 'Social capital as good culture", *Journal of the European Economic Association* **6**(2–3), 295–320.

Guiso, L., Sapienza, P. and Zingales, L. (2011), 'Civic capital as the missing link', *in* J. Benhabib, M. O. Jackson and A. Bisin, eds, 'Handbook of Social Economics', North Holland, Amsterdam.

Haleblian, J., Devers, C. E., McNamara, G., Carpenter, M. A. and Davison, R. B. (2009), 'Taking stock of what we know about mergers and acquisitions: a review and research agenda', *Journal of Management* **35**(3), 469–502.

Hamilton, G. (1995), 'Enforcement in apprenticeship contracts: were runaways a serious problem? Evidence from Montreal', *The Journal of Economic History* **55**(3), 551–574.

Hamilton, G. (1996), 'The market for Montreal apprentices: contract length and information', *Explorations in Economic History* **33**(4), 496–523.

Hamilton, G. (2000), 'The decline of apprenticeship in North America: evidence from Montreal', *The Journal of Economic History* **60**(3), 627–664.

Hanna, R. and Wang, S. Y. (2017), 'Dishonesty and selection into public service: evidence from India', *American Economic Journal: Economic Policy* **9**(3), 262–290.

Hansen, R. (1987), 'The theory of the choice of exchange medium in mergers and acquisitions', *The Journal of Business* **60**(1), 75–95.

Hansen, R. G. (1986), 'Seal-bid versus open auction: the evidence', *Economic Inquiry* **24**(1), 125–142.

Harford, J. (2005), 'What drives merger waves?', *Journal of Financial Economics* **77**(3), 529–560.

Harsanyi, J. C. (1967), 'Games with incomplete information played by "Bayesian" players, I-III. Part I. The basic model', *Management Science* **14**(3), 159–182.

Hasbrouck, J. (2009), 'Trading costs and returns for US equities: estimating effective costs from daily data', *The Journal of Finance* **64**(3), 1445–1477.

Hertzel, M. and Smith, R. (1993), 'Market discounts and shareholder gains for placing equity privately', *The Journal of Finance* **48**(2), 459–485.

Hessami, Z. (2014), 'Political corruption, public procurement, and budget composition: theory and evidence from OECD countries', *European Journal of Political Economy* **34**, 372–389.

Hsiang, S. M. (2010), 'Temperatures and cyclones strongly associated with economic production in the caribbean and central america', *Proceedings of the National Academy of Science* **107**(35), 15367–15372.

Huang, Y. and Walkling, R. (1987), 'Target abnormal returns associated with acquisition announcements: payment, acquisition form, and managerial resistance', *Journal of Financial Economics* **19**(2), 329–349.

Humphries, J. (2003), 'English apprenticeship: a neglected factor in the first Industrial Revolution', *in* P. A. David, ed., 'The economic future in historical perspective', Oxford University Press.

Kellett, J. R. (1958), 'The breakdown of gild and corporation control over the handicraft and retail trade in London', *The Economic History Review* **10**(3), 381–394.

Keshk, O. M. (2003), 'Cdsimeq: a program to implement two-stage probit least squares', *The Stata Journal* **3**(2), 1–11.

Kesner, I. F., Shapiro, D. L. and Sharma, A. (1994), 'Brokering mergers: an agency theory perspective on the role of representatives', *Academy of Management Journal* **37**(3), 703–721.

Knack, S. and Keefer, P. (1995), 'Institutions and economic performance: cross-country tests using alternative institutional measures', *Economics and Politics* **7**(3), 207–227.

Koeplin, J., Sarin, A. and Shapiro, A. (2005), 'The private company discount', *Journal of Applied Corporate Finance* **12**(4), 94–101.

Kurkchiyan, M. (2009), 'Russian legal culture: an analysis to adaptive response to an institutional transplant', *Law and Social Inquiry* **34**(2), 337–364.

Kyle, A. (1985), 'Continuous auctions and insider trading', *Econometrica* **53**(6), 1315–1335.

La Porta, R., Lopez de Silanes, F., Shleifer, A. and Vishny, R. (1999), 'The quality of government', *Journal of Law, Economics, and Organization* **15**(1), 222–279.

Laamanen, T. (2007), 'On the role of acquisition premium in acquisition research', *Strategic Management Journal* **28**(13), 1359–1369.

Lane, J. (1996), *Apprenticeship in England, 1600-1914*, University College London Press Ltd.

Lecce, G. and Ogliari, L. (2017), 'Institutional transplant and cultural proximity: evidence from nineteenth-century Prussia'. Igier Working Paper no. 598.

Leunig, T., Minns, C. and Wallis, P. (2011), 'Networks in the premodern economy: the market for London apprenticeships, 1600-1749', *The Journal of Economic History* **71**(2), 413–443.

Liberati, G. (1988), 'I demani nel Mezzogiorno continentale tra giurisprudenza e prassi', *in* A. Massafra, ed., 'Il Mezzogiorno pre-unitario. Economia, società e istituzioni', Dedalo, Bari.

Llorente, G., Michaely, R., Saar, G. and Wang, J. (2002), 'Dynamic volume-return relation of individual stocks', *Review of Financial Studies* **15**(4), 1005–1047.

Lowes, S., Nunn, N., , Robinson, J. A. and Weigel, J. (2017), 'The evolution of culture and institutions: Evidence from the Kuba kingdom.', *Econometrica* **85**(4), 1065–1091.

Ma, T.-C. (2013), 'Legal transplant, legal origin, and Antitrust effectiveness', *Journal of Competition Law and Economics* **9**(1), 65–88.

Maddala, G. S. (1983), *Limited-dependent and qualitative variables in econometrics*, Vol. 3, Cambridge University Press.

Malcomson, J. M., Maw, J. W. and McCormick, B. (2003), 'General training by firms, apprentice contracts, and public policy', *European Economic Review* **47**(2), 197–227.

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

Mauro, P. (1998), 'Corruption and the composition of government expenditure', *Journal of Public Economics* **69**(2), 263–279.

Mayshar, J., Moav, O., Neeman, Z. and Pascali, L. (2016), 'Cereals, appropriability and hierarchy'. WERPS Working Paper no. 1130.

McCormick, M., Huang, G., Zambotti, G. and Lavash, J. (2013), 'Roman road network (version 2008)'. DARMC Scholarly Data Series, Data Contribution Series 2013-5. DARMC, Center for Geographic Analysis, Harvard University, Cambridge MA.

McNamara, G. M., Haleblian, J. J. and Dykes, B. J. (2008), 'The performance implications of participating in an acquisition wave: early mover advantages, bandwagon effects, and the moderating influence of industry characteristics and acquirer tactics', *Academy of Management Journal* **51**(1), 113–130.

Melillo, M. (1959), 'Intorno alle probabili sedi originarie delle colonie francoprovenzali di Celle e Faeto', *Revue de linguistique romane* **23**, 1–34.

Mennonna, A. R. (1987), *I dialetti gallitalici della Lucania*, Congedo Editore, Galatina.

Miller, E. (1977), 'Risk, uncertainty, and divergence of opinion', *The Journal of Finance* **32**(4), 1151–1168.

Ministero per i Beni e le Attività Culturali (1999-2001), *Guida alle fonti per la storia del brigantaggio postunitario conservate negli Archivi di Stato, voll. I-III*, Ufficio Centrale per i Beni Archivistici, Roma.

Minns, C. and Wallis, P. (2012), 'Rules and reality: quantifying the practice of apprenticeship in early modern England', *The Economic History Review* **65**(2), 556–579.

Minns, C. and Wallis, P. (2013), 'The price of human capital in a pre-industrial economy: premiums and apprenticeship contracts in 18th-century England', *Explorations in Economic History* **50**(3), 335–350.

Mo, P. H. (2001), 'Corruption and economic growth', *Journal of Comparative Economics* **29**(1), 66–79.

Moeller, S., Schlingemann, F. and Stultz, R. (2007), 'How do diversity of opinion and information asymmetry affect acquirer returns?', *Review of Financial Studies* **20**(5), 2047–2978.

Molfese, F. (1964), *Storia del brigantaggio dopo l'unità*, Feltrinelli, Milano.

Montanari, D. (1999), *Monti di pietà e presenza ebraica in Italia (secoli XV-XVIII)*, Bulzoni, Roma.

Murphy, K., Shleifer, A. and Vishny, R. W. (1991), 'The allocation of talent: implications for growth', *Quarterly Journal of Economics* **106**(2), 503–530.

Myers, S. and Majluf, N. (1984), 'Corporate financing and investment decisions when firms have information that investors do not have', *Journal of Financial Economics* **13**(2), 187–221.

Officer, M. (2003), 'Termination fees in mergers and acquisitions', *Journal of Financial Economics* **69**(3), 431–467.

Officer, M. (2007), 'The price of corporate liquidity: acquisition discounts for unlisted targets', *Journal of Financial Economics* **83**(3), 571–598.

Officer, M., Poulsen, A. and Stegemoeller, M. (2009), 'Target-firm information asymmetry and acquirer returns', *Review of Finance* **13**(3), 467–493.

Ogilvie, S. (2004), 'Guilds, efficiency and social capital: evidence from German proto-industry', *The Economic History Review* **57**(2), 286–333.

Olken, B. A. (2009), 'Corruption perceptions vs. corruption reality', *Journal of Public Economics* **93**(7), 950–964.

Olken, B. A. and Barron, P. (2009), 'The simple economics of extortion: evidence from trucking in Aceh', *Journal of Political Economy* **117**(3), 417–452.

Papageorgiou, T. (2014), 'Learning your comparative advantages', *The Review of Economic Studies* **81**(3), 1263–1295.

Pappalardo, F. (2014), *Dal banditismo al brigantaggio. La resistenza allo Stato moderno nel Mezzogiorno d'Italia*, D'Ettoris, Crotone.

Pascali, L. (2016), 'Banks and development: Jewish communities in the Italian Renaissance and current economic performance', *Review of Economics and Statistics* **98**(1), 140–158.

Pastor, L. and Stambaugh, R. (2003), 'Liquidity risk and expected stock return', *Journal of Political Economy* **111**(3), 642–685.

Pfister, M. (1991), 'Gerhard Rohlfs e le colonie gallo-italiche nella Basilicata', *in* N. De Blasi, P. Di Giovine and F. Fanciullo, eds, 'Le parlate lucane e la dialettologia italiana (Studi in memoria di Gerhard Rohlfs)', Congedo, Galatina.

Putnam, R. D. (1993), *Making democracy work. Civic traditions in modern Italy*, Princeton University Press.

Reuer, J. J., Tong, T. W. and Wu, C.-W. (2012), 'A signaling theory of acquisition premiums: evidence from IPO targets', *Academy of Management Journal* **55**(3), 667–683.

Rhodes-Kropf, M. and Viswanathan, S. (2004), 'Market valuation and merger waves', *The Journal of Finance* **59**(6), 2685–2718.

Rizzica, L. (2016), 'Why go public? A study of the individual determinants of public sector employment choice'. Bank of Italy, Occasional Paper no. 343.

Rizzica, L. and Tonello, M. (2015), 'Exposure to media and corruption perceptions'. Bank of Italy, Working Paper no. 1043.

Rohlfs, G. (1972), *Studi e ricerche su lingua e dialetti d'Italia*, Sansoni Editore, Firenze.

Rohlfs, G. (1985), *Dizionario storico dei cognomi in Lucania*, Longo Editore, Ravenna.

Rohlfs, G. (1988), *Studi linguistici sulla Lucania e sul Cilento*, Congedo Editore, Galatina.

Roll, R. (1984), 'A simple implicit measure of the effective bid-ask spread in an efficient market', *The Journal of Finance* **39**(4), 1127–1139.

Russo, S. (1988), 'Distribuzione della proprietà, stratificazione e mobilità sociale a Cerignola nell'Ottocento', *in* A. Massafra, ed., 'Il Mezzogiorno pre-unitario. Economia, società e istituzioni', Dedalo, Bari.

Scarangello, A. A. (1964), *Progress and Trends in Italian Education*, U.S. Office of Education Bulletin no. 21, Washington, D.C.: U.S. Department of Health, Education, and Welfare.

Schwarz, L. D. (1987), 'London apprentices in the seventeenth century: some problems', *Local Population Studies* **38**, 18–22.

Schwarz, L. D. (1992), *London in the age of industrialisation. Entrepreneurs, labour force and living conditions, 1700-1850*, Cambridge University Press.

Schwert, G. (2002), 'Hostility in takeovers: in the eyes of the beholder?', *The Journal of Finance* **55**(6), 2599–2640.

Servaes, H. (1991), 'Tobin's Q and the gains from takeovers', *The Journal of Finance* **46**(1), 409–419.

Shaw-Taylor, L. and Wrigley, E. A. (2008), 'The occupational structure of England c. 1750-1871: a preliminary report'. Cambridge Group for the History of Population and Social Structure.

Shleifer, A. and Vishny, R. W. (1993), 'Corruption', *Quarterly Journal of Economics* **108**(3), 599–617.

Smith, R. L. (1987), 'The choice of issuance procedure and the cost of competitive and negotiated underwriting: an examination of the impact of Rule 50', *The Journal of Finance* **42**(3), 703–720.

Smith, S. R. (1973), 'The social and geographical origins of the London apprentices: 1630-1660', *Guildhall Miscellany* **4**, 195–206.

Spolaore, E. and Wacziarg, R. (2016*a*), 'Ancestry, language and culture', *in* V. Ginsburgh and S. Weber, eds, 'The Palgrave Handbook of Economics and Language', Palgrave Macmillan, London.

Spolaore, E. and Wacziarg, R. (2016*b*), 'Fertility and modernity'. UCLA CCPR Population Working Paper.

Stultz, R. M. (1988), 'Managerial control of voting rights: financing policies and the market for corporate control', *Journal of Financial Economics* **20**(1), 25–54.

Tabellini, G. (2008), 'The scope of cooperation: values and incentives', *Quarterly Journal of Economics* **123**(3), 905–950.

Tabellini, G. (2010), 'Culture and institutions: economic development in the regions of Europe', *Journal of the European Economic Association* **8**(4), 677–716.

Tabellini, G. and Greif, A. (2010), 'Cultural and institutional bifurcation: China and Europe compared', *American Economic Review* **100**(2), 135–140.

Tadelis, S. and Rangel, A. (2001), 'Apprenticeships: human capital and competitive signaling in a dynamic labor market'. SIEPR Discussion Paper no. 01-03.

Toso, F. (2002), 'Il galloitalico di Lucania: contributo alla precisazione dell'area di origine', *in* G. Holtus and J. Kramer, eds, 'Ex traditione inovatio. Max Pfister und die Schweizer Romanistik', Niemayer, Tübingen.

Travlos, N. G. (1987), 'Corporate takeover bids, method of payment, and bidding firms' stock returns', *The Journal of Finance* **42**(4), 943–963.

Uslaner, E. M. (2004), 'Trust and corruption', *in* J. G. Lambsdorf, M. Taube and M. Schramm, eds, 'The New Institutional Economics of corruption', Routledge, London.

Vegezzi Ruscalla, G. (1862), 'Colonia piemontese in Calabria: studio etnografico', *Rivista contemporanea* **28**.

Villani, P. (1964), *La vendita dei beni dello Stato nel Regno di Napoli (1806-1815)*, Banca Commerciale Italiana, Milano.

Voigtländer, N. and Voth, H.-J. (2012), 'Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany', *Quarterly Journal of Economics* **127**(3), 1339–1392.

Wallis, P. (2008), 'Apprenticeship and training in premodern England', *The Journal of Economic History* **68**(3), 832–861.

Wallis, P., Webb, C. and Minns, C. (2010), 'Leaving home and entering service: the age of apprenticeship in early modern London', *Continuity and change* **25**(3), 377–404.

Wrigley, E. A., Schofield, R. S. and Schofield, R. (1989), *The population history of England 1541-1871*, Cambridge University Press.

Yarbrough, A. (1980), 'Geographical and social origins of Bristol apprentices, 1542–1565', *Transactions of the Bristol and Gloucestershire Archaeological Society* **98**, 113–130.